

THE
BRITISH JOURNAL
OF
PSYCHOLOGY

Statistical Section

EDITED BY
CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF

A. C. AITKEN	L. S. PENROSE
M. S. BARTLETT	J. FRASER ROBERTS
W. G. EMMETT	A. RODGER
M. G. KENDALL	B. BABINGTON SMITH
D. N. LAWLEY	W. STEPHENSON
P. E. VERNON	

Managing Editor : H. G. MAULE

Volume II
1949

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4.

Digitized by the Internet Archive
in 2026

ON ESTIMATING OBLIQUE FACTORS

By GODFREY THOMSON

Moray House, University of Edinburgh

I. *Object of this Note.* II. *Shortened Method of Estimating Oblique Factors.*

I. OBJECT OF THIS NOTE

The main object is to point out how Ledermann's shortened method of estimating mental factors by regression (1 and 2) must be modified to include the case of oblique factors.

If R is the matrix of correlations of the tests, with unities in its diagonal cells, and if the correlations of the n tests with the r common factors are given by the $n \times r$ matrix M_0 , while the correlations of the n tests with the n specifics are given by the diagonal matrix M_1 , then the regression coefficients to be applied to the standardized test scores in order to estimate a man's factors from his test scores are $M_0' R^{-1}$. As this involves finding the reciprocal of a large $n \times n$ matrix, Ledermann devised the above-mentioned short method, by which the same regression coefficients are given by $(I + J)^{-1} M_0' M_1^{-2}$, where I is the unit matrix and $J = M_0' M_1^{-2} M_0$. Since $(I + J)$ is only of order r , this method only involves finding the reciprocal of a much smaller matrix than R , and is especially economical when the tests are much more numerous than the common factors.

In the case of orthogonal factors (and when Ledermann wrote no others were in general use) the matrix M_0 of correlations between tests and factors is what is commonly called the matrix of loadings (or saturations). When the factors are orthogonal these loadings are *both* the correlations with the tests, *and also* the coefficients in the linear specification equations expressing the test scores in terms of the factor values.

When, however, oblique factors are used, as in L. L. Thurstone's system, this is no longer the case; the coefficients in the specification equations are no longer identical with the correlations between the factors and the tests. The former are the oblique co-ordinates of the test-point on the factor-axes, the latter are its projections on to those axes.

In Thurstone's system certain axes called Reference Vectors are used; but these are only means to an end, which is to identify the Oblique Factors which correspond with psychological entities like memory, word comprehension, perception, and so on. It is the factors we want to estimate in a man, and we need therefore the matrix of their correlations with the tests.

Let F_0 be the $n \times r$ matrix of centroid loadings, and F_1 the diagonal matrix of specific loadings, so that

$$F_0 F_0' + F_1^2 = R.$$

Then it can be shown (3, p. 375) that the required correlations of the tests with Thurstone's oblique factors are

$$F_0 (\Lambda')^{-1} D,$$

where Λ is the matrix required to rotate the centroid factors into simple structure on the reference vectors and D is a diagonal matrix whose purpose is to normalize the columns of $(\Lambda')^{-1}$ to make them into direction cosines.

The regression coefficients, therefore, for estimating the oblique factors from the test scores by the long method are

$$\{F_0(\Lambda')^{-1}D\}'R^{-1}.$$

II. SHORTENED METHOD OF ESTIMATING OBLIQUE FACTORS

Before we can parallel Ledermann's short formula, however, we must first note that when axes are oblique the formula

$$F_0F_0' + F_1^2 = R,$$

true for orthogonal axes, requires modification. Its analogue now is

$$\text{Pattern} \times \text{Transpose of Structure} + F_1^2 = R,$$

where 'pattern' means the matrix of specification coefficients (oblique co-ordinates) and 'structure' means the matrix of correlations (projections on to the axes, or cosines of the angles). The latter is the matrix $F_0(\Lambda')^{-1}D$ mentioned above. The pattern (v. 3, p. 375) is $F_0\Lambda D^{-1}$. We have indeed

$$\begin{aligned} (F_0\Lambda D^{-1})\{F_0(\Lambda')^{-1}D\}' + F_1^2 \\ = F_0\Lambda D^{-1}D\Lambda^{-1}F_0' + F_1^2 \\ = F_0F_0' + F_1^2 = R. \end{aligned}$$

We can now parallel Ledermann's deduction of his short method as follows :

$$\begin{aligned} \{F_0(\Lambda')^{-1}D\}'F_1^2R &= \{F_0(\Lambda')^{-1}D\}'F_1^2[(F_0\Lambda D^{-1})\{F_0(\Lambda')^{-1}D\}' + F_1^2] \\ &= [\{F_0(\Lambda')^{-1}D\}'F_1^2(F_0\Lambda D^{-1}) + I]\{F_0(\Lambda')^{-1}D\}' \\ &= (J + I)\{F_0(\Lambda')^{-1}D\}', \end{aligned}$$

whence, premultiplying by $(I + J)^{-1}$ and postmultiplying by R^{-1} , we obtain

$$\begin{aligned} (I + J)^{-1}\{F_0(\Lambda')^{-1}D\}'F_1^2 &= \{F_0(\Lambda')^{-1}D\}'R^{-1} \\ &= \text{the required regression coefficients.} \end{aligned}$$

Here $J = \{F_0(\Lambda')^{-1}D\}'F_1^2(F_0\Lambda D^{-1})$

in place of $J = M_0'M_1^{-2}M_0$ in the orthogonal case considered by Ledermann. The matrices whose reciprocals have to be calculated are here, as there, of order $r \times r$ only. But the calculation is still a complex one, as indeed are all calculations concerning oblique factors—a serious practical drawback.

REFERENCES

1. Ledermann, W. (1938). A shortened method of estimation of mental factors by regression, *Nature*, CXLI, p. 650.
2. Ledermann, W. (1939). A shortened method of estimation of mental factors by regression, *Psychometrika*, IV, pp. 109-16.
3. Thomson, Godfrey H. (1948). *The Factorial Analysis of Human Ability* (3rd edition). London : University of London Press.

EVIDENCE OF A SPACE FACTOR AT 11+ AND EARLIER

By W. G. EMMETT

Education Department, Moray House, University of Edinburgh

I. Introduction. II. Implications of a Space Factor. III. Sex Differences in Spatial Judgment. IV. Lawley's Method of Factor Analysis. V. The Moray House 11+ Enquiry. VI. Margaret Mellone's 7+ Battery. VII. Drew's 1947 Enquiry. VIII. El Koussy's 1935 Enquiry. IX. Slater's 1943 Enquiry. X. Summary.

I. INTRODUCTION

The present enquiry started with the construction of a test, known as Moray House Space Test 1, which it was hoped would select those children at the age of 11+ who could profit most from secondary school courses requiring the special ability to perceive spatial relations. The work was initiated in 1944 by Professor Godfrey Thomson assisted by Mr. John H. Gray and Mr. J. Y. Erskine, and later devolved on the present writer in conjunction with Mr. J. T. Bain and Mr. L. F. Mills. Mr. Bain completed the construction of the test in 1946; in the following year Mr. Mills provided the material for the intercorrelations of a battery of tests, including the space test, and carried out a preliminary factorial analysis. Further statistical work by Professor Thomson and the writer showed the existence of a significant group factor associated with the space test.

Indications of a space factor amongst 7+ boys had previously been found in a Moray House Pictorial Intelligence Test, M.H.T.Pic. 1, by Margaret Mellone (10). Patrick Slater (14, 15), however, could find no evidence of any such factor at either 11+ or 13+. In consequence the writer made factorial analyses, by D. N. Lawley's Maximum Likelihood Method, of the correlations obtained by the above-mentioned and other authors. In every case a significant group factor associated with one or more of the spatial tests was found. Additional evidence of this factor was found from sex differences in spatial and verbal tests.

The statistical data which follow are the result of calculations to four or more significant figures. A few seeming inconsistencies in the additions may have arisen from rounding off to the nearest third significant figure. Factor loadings have been expressed in terms of orthogonal factors, as these have been used by those workers whose data are here under review, and may perhaps be more easily interpreted than oblique factors.

II. IMPLICATIONS OF A SPACE FACTOR

The existence of a space factor implies individual differences in the power to perceive spatial relations, this differentiation occurring independently of that in general intelligence and special abilities such as verbal, numerical, memory, word-fluency, and the like. It is probable that persons highly endowed with the space factor will

achieve success in subjects such as draughtsmanship, woodwork, metal work, needlework, modelling, and most science subjects, particularly biology and surgery. There is already much evidence that space tests given in the late teens are predictive of 'technical proficiency': cf. F. Holliday (3, 4) and C. W. Shuttleworth (13).¹ If, therefore, it can be shown that such tests given in early school life have the same factorial composition as similar tests given later, there will be good grounds for hoping that they will select the younger children as successfully as they do the older. But the issue is only to be decided finally by correlation with later performance.

It is sometimes asserted by psychologists that the abilities of young children are less specialized than those of older (cf. Burt (18), p. 132), and that in particular the special ability to perceive spatial relations, that is, the spatial factor, does not develop until puberty or after (cf. Slater (15)). If this were so, then the children selected at 11+ by a spatial test would not be the same as those selected later, for the spatial ability would develop independently of those abilities by which selection at 11+ was made. On the other hand, if the spatial ability exists at 11+, an order of merit in spatial tests is likely to be stable over a period, if one argues from the stability of performances in intelligence, verbal, and other tests, which measure abilities well defined at 11+.

III. SEX DIFFERENCES IN SPATIAL JUDGMENT

The superiority of boys to girls in certain spatial and performance tests has been established beyond all doubt. When the same groups of boys and girls, both forming a complete year-group in an educational area, are compared in verbal ability, either by a verbal intelligence test or an English test, the girls are at least on the same level as the boys, but more often significantly superior, especially since the war, say, since 1945.

Since there is no reason to think that the sexes differ in average general intelligence, it can be argued that the girls' superiority in verbal ability arises from a mental function independent of general intelligence, in other words, from a verbal factor; and similarly that the boys' superiority in space tests derives from a spatial factor. The argument is, of course, applicable to any two groups, irrespective of sex, which while equal in intelligence show a difference in one or more abilities.

In Table I we give statistics of some observed sex differences. In almost every case the differences are highly significant.

The sex differences in rows 1 and 2 of Table I were observed by Margaret Mellone when constructing a picture intelligence test for 7-year olds; those in row 3 resulted from a pilot survey in Halifax in 1932 in connexion with the Scottish Mental Survey of that year. In both cases the differences were so great as to preclude the use of the tests for a mixed group of children.

In rows 4 and 5 the scores in the Moray House space test were derived from two complete year-groups in industrial county boroughs (Areas G and AF in Moray House records). The test items were divided into two groups, those demanding spatial judgment in two dimensions and in three dimensions. The sex difference persists for each category of item and is highly significant. A greater sex difference is observed for the three-dimensional items than for the two-dimensional, and though this difference in boys' superiority

¹ Shuttleworth gives useful references at the end of his article.

TABLE I. SEX DIFFERENCES IN SPATIAL AND VERBAL ABILITY

Test	No. of items	Age range	Boys		Girls		Diff. B-G	Reference
			N	Mean	N	Mean		
A. Spatial Ability								
1. Cube Counting	15	6 : 7 to 7 : 6	218	61.4%	196	46.5%	14.9%	Mellone (10)
2. Mazes	15	6 : 7 to 7 : 6	218	56.3%	196	24.0%	32.3%	"
3. Cube Counting	18	10 : 8 to 11 : 7	660	55.2%	618	42.3%	12.9%	S.C.R.E. (12)
M.H.T. Space 1 :								
4. 2 dimensions	52	10 : 3 to 11 : 3	1,856	20.85	1,747	19.53	1.32	Moray House records, 1947
5. 3 dimensions	48	" "	1,856	22.87	1,747	20.67	2.20	
6. Performance test battery	—	8 : 11 to 11 : 9	443	56.32	430	52.55	3.77	MacMeeken (9)
7. W. P. Alexander's performance scale	—	11 : 0	—	112	—	100	12	Table of norms
8. Space Test, M.H.T. Sp. 1	100	10 : 3 to 11 : 6	4,405	45.21	4,349	42.47	2.74	Moray House records, 1947
B. Verbal Ability								
9. Verbal Intelligence Tests (Moray House)	100	10 : 3 to 11 : 6	4,512	45.76	4,413	47.88	-2.12	Moray House records, 1947
10. Verbal Intelligence, M.H.T. 21, 1935	100	10 : 10 to 11 : 9	3,450	49.90	3,439	51.91	-2.01	Moray House records, Area A
11. Verbal Intelligence, M.H.T. 21, 1946	100	10 : 3 to 11 : 2	2,669	43.23	2,642	49.20	-5.97	
12. Verbal Intelligence, M.H.T. 28, 1938	100	10 : 2 to 11 : 1	7,065	40.28	6,705	40.63	-0.35	Moray House records, Area M
13. Verbal Intelligence, M.H.T. 28, 1947	100	10 : 2 to 11 : 1	6,615	38.57	6,380	41.33	-2.76	

is not significant ($P = 0.10$) its trend supports evidence from factor analysis, given later, that three-dimensional items have a higher loading in the space factor than two-dimensional.

The space test and verbal intelligence test scores in rows 8 and 9 came from complete year-groups in six educational districts, the same children taking both tests.

The figures in rows 10 to 13, though not strictly relevant, are of interest as showing an increased superiority of girls in verbal intelligence of recent years : they refer to complete year-groups of children in the two areas which took the same test at widely different times. A similar superiority has been found in most other districts. Since the superiority is even greater in English tests, but does not occur in arithmetic tests, it presumably arises from a greater verballity rather than from a genuine increase in general intelligence relative to boys. More recently we have found that the gap between boys and girls is closing up.

IV. LAWLEY'S METHOD OF FACTOR ANALYSIS

Any success achieved in the present enquiry in establishing a space factor is due entirely to the use of D. N. Lawley's method of factor analysis (6, 7, 8). The method when applied to large samples, say, of 200 cases or more, makes the most efficient use of the statistical information, minimizing the error variance due to sampling. It appears to give greater weight to those variables with the higher correlations, about which therefore most is known, whereas Thurstone's centroid method gives

equal weight to all variables. Equally or perhaps more important, it provides for large samples a satisfactory test of the significance of the residual correlations after removal of factors whose significance has previously been established. Most, if not all, tests of significance hitherto in use are based on approximations not always justified, and it is possible that a factor on the borderline of significance may be accepted as significant or otherwise on inadequate grounds.

Quite recently Dr. Lawley has devised a test of significance of the factor loadings themselves before rotation ; a paper on the subject is due for publication shortly. Meanwhile he has very kindly placed his formulæ at our disposal, and we have used them in cases where the residual correlations of the whole battery were not highly significant, but where a particular test was suspected of having a significant loading in a new factor.

The factors resulting from the method of analysis are orthogonal, and reference axes may be rotated by orthogonal or oblique transformations according to taste.

An example of the practical application of this method has been published (Lawley (8), and Thomson (17), Chap. 21) ; it relates to a two-factor analysis and perhaps requires a little amplification when more factors are in question. This the present writer, with Dr. Lawley's assent, hopes to give soon along with checks at each stage of calculation.

The calculations are laborious. A nine-variable analysis into three factors may take 30 working hours with the aid of a calculating machine, but as this time is short compared with that taken in collecting the data and computing the correlations it is certainly well spent. For more variables the time required is roughly proportional to the square of the number of variables. The work is almost prohibitive for more than 15 variables, if four or more factors are expected.

V. THE MORAY HOUSE 11+ ENQUIRY

As stated in Section I, this enquiry revolved round a newly constructed space test, whose value in predicting success in certain technical school subjects it was proposed to investigate. The greater part of the work in constructing the test fell upon Mr. J. T. Bain. The test is a paper-pencil test of 100 items selected as the most discriminating from about 200 draft items ; 52 of the items involve two-dimensional spatial judgment and 48 three-dimensional. A mean test score of between 40 and 45 for a year-group of mean age about 11 : 0 was obtained by adjusting the time allowed for its five separately timed sections. The standard deviations of raw scores for five complete year-groups of 11+ boys averaged 23·4 ($N=3,710$) and of 11+ girls ($N=3,665$) in the same areas 21·2. The reliability coefficient of the test was obtained from a random sample of 200 scripts taken from a complete 11+ year-group. By correlating odd and even items and applying the Spearman-Brown formula, the coefficient was 0·965. From answer pattern data on the same sample the Kuder-Richardson formula gave a coefficient of 0·959. The test is thus highly discriminative.

For the factorial analysis a battery of nine tests was put together by Mr. L. F. Mills. Scores in four of the variables were already available from Moray House tests used in the Edinburgh Qualifying Examination a short time previously : these were intelligence quotients, English quotients, and raw scores in mechanical and

problem arithmetic. The space test was divided according to the two- and the three-dimensional items, giving a further two variables. The battery was completed by the Raven 1938 matrix test given without time limit, the non-verbal intelligence test 70/23 of the National Institute of Industrial Psychology by Patrick Slater, and a similar test from the National Foundation for Educational Research in England and Wales by J. W. Jenkins. These three tests were included as it was thought possible that some spatial ability might enter into the solution of their problems. A larger battery was precluded by limitations of time.

Table II gives certain statistics of the nine variables.

TABLE II. EXPERIMENTAL DATA IN THE MORAY HOUSE 11+ ENQUIRY, 1946-1948

Tests given to 178 boys ; last year pupils in Edinburgh primary schools, ages 11 : 0 to 12 : 11, 94% between 11 : 8 and 12 : 9, mean age 12 y. 1.6 m. Intelligence quotient : mean, 99.44 ; s.d., 11.66. English quotient : mean, 102.41 ; s.d., 12.87.

Tests	No. of items	Mean score	s.d.	Relative s.d. of raw scores *
1. Verbal Intelligence, M.H.T. 41	100	—	—	25.71
2. English, M.H.E. 17	120	—	—	23.95
3. Arithmetic, Mechanical, M.H.A. 17 (i)	50	30.09	9.14	22.37
4. Arithmetic, Problems, M.H.A. 17 (ii)	50	31.69	10.97	26.84
5. Matrix, Raven 1938	60	37.42	8.16	16.64
6. Non-verb. Intelligence, N.I.I.P. 70/23, Slater	53	21.23	5.86	13.53
7. Non-verb. Intelligence, Nat. Found., Jenkins	85	53.10	13.61	19.59
8. Space, 2 dimensions	52	28.57	11.09	26.09
9. Space, 3 dimensions	48	22.59	8.75	22.30

* The relative standard deviations in the last column have been calculated on the basis of a 100-item test given to a complete 11+ year-group.

The scores in tests 1 and 2 were obtained as standard scores with mean of 100 and standard deviation of 15 for an unselected population. They are henceforward termed 'quotients,' though they are not strictly such. For the other variables raw scores and their standard deviations are recorded. In the last column of the table are shown the standard deviations of raw scores which would have been obtained with a 100-item test given to a complete population with standard deviation of quotient of 15. The figures for tests 1 and 2 were obtained from direct measurement over several complete 11+ year-groups. For the other tests corrections by direct proportionality have been applied. The standard deviation of quotient of the sample has been taken as the mean of the intelligence and English quotients, viz., 12.26. The 'relative s.d.,' for example, for test 6 is then given by $5.86 \times \frac{15}{12.26} \times \frac{100}{53} = 13.53$. In a search for new factors it is important to use tests with high

discriminatory power and a sample of the population as nearly representative as possible, since the resulting high correlations make it easier to establish the significance of those factors which appear. We were fortunate in having at our disposal tests specially designed for 11+ children ; the low relative standard deviations of tests 5 and 6 can probably be ascribed to their having been designed to cover a wider age range than Moray House tests.

After Mr. Mills had administered tests 5 to 9 in Edinburgh primary schools with all necessary precautions, he calculated the correlations of the nine variables amongst

themselves and with age, whence a set of correlations freed from the effect of age was obtained. These are given in Table III.

TABLE III. CORRELATIONS FOR THE MORAY HOUSE 11+ ENQUIRY

Tests	1	2	3	4	5	6	7	8	9
Verbal Intelligence .. 1	—	.794	.725	.763	.643	.587	.675	.668	.395
English .. 2	.794	—	.727	.725	.512	.456	.569	.547	.301
Arithmetic, Mechanical .. 3	.725	.727	—	.759	.344	.333	.411	.379	.196
Arithmetic, Problems .. 4	.763	.725	.759	—	.477	.493	.549	.616	.349
Matrix, Raven .. 5	.643	.512	.344	.477	—	.514	.731	.622	.413
Non-verb. Intell., N.I.I.P. 6	.587	.456	.333	.493	.514	—	.601	.630	.394
Non-verb. Intell., Jenkins 7	.675	.569	.411	.549	.731	.601	—	.627	.470
Space, 2 dimensions .. 8	.668	.547	.379	.616	.622	.630	.627	—	.621
Space, 3 dimensions .. 9	.395	.301	.196	.349	.413	.394	.470	.621	—

Mr. Mills then made a preliminary centroid analysis of these correlations, and in the time at his disposal found two significant factors. The analysis was then pursued first by Professor Godfrey Thomson and Mr. Swanson and later by the present writer, using Lawley's method of maximum likelihood. Four significant factors were found, three being evaluated by Lawley's method and the fourth by centroid analysis of the resulting residual correlations. The fourth factor had appreciable loadings¹ in tests 5, 7, and 9, and of opposite sign in tests 6 and 8; as it was difficult to interpret psychologically it was not taken into further account. Axes were rotated orthogonally: I_0 and II_0 so that I_1 passed through test 8, I_1 and III_0 so that I_2 passed through test 7, and finally II_1 and III_1 so that II_2 passed through test 3. Table IV gives the loadings before and after rotation.

TABLE IV. FACTOR LOADINGS FOR THE MORAY HOUSE 11+ BATTERY

Tests	Before rotation				After rotation		
	I_0	II_0	III_0	h^2	I_2	II_2	III_2
Verbal Intelligence .. 1	.920	.055	-.082	.855	.785	.489	.010
English .. 2	.833	.200	-.050	.737	.638	.574	-.016
Arithmetic, Mechanical .. 3	.765	.487	.070	.828	.421	.807	.000
Arithmetic, Problems .. 4	.844	.190	.131	.765	.606	.611	.160
Matrix, Raven .. 5	.695	-.359	-.273	.687	.826	-.012	-.066
Non-verb. Intell., N.I.I.P. .. 6	.641	-.295	-.037	.499	.691	.066	.131
Non-verb. Intell., Jenkins .. 7	.749	-.324	-.212	.712	.841	.058	-.013
Space, 2 dimensions .. 8	.758	-.396	.186	.766	.781	.089	.385
Space, 3 dimensions .. 9	.494	-.464	.372	.598	.538	-.060	.552
Variance, $S (I^2)$	5.111	1.009	.325	6.445	4.333	1.612	.501
Per cent. of total variance ..	56.79	11.21	3.61	71.61	48.15	17.92	5.57

¹ We use the term 'loading' rather than that confusing term 'saturation.' The latter implies an analogy between a test and a liquid solution, and suggests that a test may have any factor content (or loading) from zero to saturation point according to circumstances—which is absurd.

The first rotated factor may be taken as g . The second factor appears only in tests 1, 2, 3, and 4, tests in which verbal ability and schooling dominate. Although tests 3 and 4 are arithmetic tests, it may well be that successful performance is largely determined by a ready understanding of the verbal medium in which the subject is taught. We have a similar conjunction of reading and arithmetic in the factorial analysis of Mellone's 7+ battery given later. The identity of the factor is not important for our present purpose; we may accept it as v tentatively.

The third factor is associated almost entirely with the two sections of the space test, with a higher loading for three-dimensional perception. The sex differences already quoted suggest a differentiation between two- and three-dimensional perception, and it may be that further investigation will effect resolution into two independent spatial factors. This third factor we may for the time being take as Koussy's space factor k .

The loadings in the space test as a whole have been found from the correlation of each factor with the sum of tests 8 and 9, each test weighted according to the standard deviation of its raw scores. These loadings along with those of the Moray House verbal intelligence test are given below.

	g	v	k	h^2
M.H. Space Test	.747	.026	.509	.818
M.H. Verbal Intelligence Test	.785	.489	.010	.855

The space test thus appears almost as good a measure of $(g+k)$ as the verbal test is of $(g+v)$. Its communality is high, and would naturally be still higher with a representative sample of children with a standard deviation of I.Q. at 15 instead of 11.66 as in the experimental sample. Incidentally, the absence of any verbal factor in the space test indicates that the verbal instructions of the test were sufficiently within the grasp of all.

VI. MARGARET MELLONE'S 7+ BATTERY

In 1943 Dr. Mellone (10) carried out a factor analysis of the sub-tests of an experimental 7+ picture intelligence test then in course of construction. She found a significant third factor for boys, but none for girls. Lawley's maximum likelihood method of analysis was used, but in consequence of certain short cuts some doubt attached to the results. Accordingly Mellone's correlations were analysed *de novo* by Lawley's method. Her data were obtained from 218 boys and 196 girls, aged 6:7 to 7:6, first-year pupils in seven Edinburgh primary schools. The mean reading quotients (Vernon) were 103.2 for boys and 106.5 for girls; the mean arithmetic quotients (Ballard) were 108.4 for boys and 110.7 for girls.

Analysis of the boys' correlations gave a highly significant set of second residuals ($P=0.0004$). For the girls the second residuals were significant only at $P=0.057$.

The factor loadings before and after orthogonal rotation are given in Table V.

TABLE V. FACTOR ANALYSIS OF MELLONE'S BATTERY

A. Loadings before rotation

Tests	218 Boys					196 Girls			
	I ₀	II ₀	III ₀	h ²		I ₀	II ₀	III ₀	h ²
Substitution .. 1	·396	·209	·268	·273	1	·343	·119	·101	·142
Absurdities .. 2	·608	·226	·272	·495	2	·613	·265	—·106	·458
Memory Span .. 3	·458	·072	·317	·315	3	·443	·090	—·154	·228
XO Series .. 4	·648	·041	—·108	·433	4	·635	·078	—·082	·417
Analogies .. 5	·558	·165	·030	·339	5	·631	·124	·142	·434
Cube Counting .. 6	·400	—·104	—·277	·247	6	·286	·054	—·185	·119
Directions .. 7	·450	—·027	—·035	·205	7	·568	—·086	·068	·334
Doesn't Belong .. 8	·622	·138	·052	·408	8	·686	·093	—·078	·486
Always Has .. 9	·572	·007	—·133	·345	9	·585	·162	·257	·435
Completion .. 10	·689	·020	—·179	·506	10	·649	·131	—·059	·442
Sequence .. 11	·669	·036	·031	·450	11	·632	·105	—·151	·433
Mirror Images .. 12	·740	·146	—·157	·593	12	·665	·098	·151	·474
Reading Quotient .. 14	·439	—·564	·083	·517	14	·398	—·637	·145	·585
Arith. Quotient .. 15	·447	—·571	·092	·535	15	·410	—·649	—·145	·610
Variance, <i>S</i> (<i>I</i> ²) ..	4·406	·827	·430	5·662		4·299	1·023	·274	5·596
Per cent. of total variance	31·47	5·91	3·07	40·45		30·70	7·31	1·96	39·97

Test No. 13 (mazes) was omitted from the analysis as it was not administered to the girls.

B. Loadings after rotation

The variables have been rearranged approximately in ascending order of magnitude of the third factor for boys.

Tests		Boys				Girls				
		I ₂	II ₂	III ₂	h ²	I ₂	II ₁	III ₁	h ²	
Reading Quotient	.. 14	·312	·648	·005	·518	14	·302	·703	·000	·585
Arith. Quotient	.. 15	·323	·656	·000	·534	15	·170	·717	·259	·610
Substitution	.. 1	·506	—·124	·042	·273	1	·364	—·049	·085	·142
Memory Span	.. 3	·561	·018	—·003	·315	3	·322	·000	·352	·228
Absurdities	.. 2	·682	—·080	·152	·495	2	·522	—·139	·408	·458
Directions	.. 7	·341	·158	·252	·205	7	·506	·196	·200	·334
Analogies	.. 5	·495	—·003	·307	·339	5	·632	·003	·185	·434
Doesn't Belong	.. 8	·554	·038	·315	·408	8	·568	·044	·401	·486
Sequence	.. 11	·565	·151	·329	·450	11	·488	·022	·440	·433
Cube Counting	.. 6	·151	·250	·403	·247	6	·166	·004	·302	·119
Always Has	.. 9	·389	·174	·404	·345	9	·655	—·043	·067	·435
XO Series	.. 4	·470	·159	·432	·433	4	·520	·048	·380	·417
Missing Part	.. 10	·460	·201	·505	·506	10	·552	·000	·371	·442
Mirror Images	.. 12	·534	·092	·547	·593	12	·660	·035	·191	·474
Variance, <i>S</i> (<i>I</i> ²)	..	3·117	1·088	1·457	5·662		3·312	1·076	1·207	5·596
Per cent. of total variance		22·27	7·77	10·41	40·45		23·66	7·69	8·62	39·97

For boys, axes I₀ and II₀ were rotated so that I₁ passed through test 12; rotation through a greater angle would have given unreasonably small loadings of I₁ in tests 14 and 15. Tests 1 and 2 are thereby left with small negative loadings in II₁. Axes I₁ and III₀ were then rotated so that I₂ passed through test 3, and finally II₁ and III₁ so that II₂ passed through Test 15, thus reducing the negative loadings in tests 1 and 2.

For girls, axes I_0 and II_0 were rotated so that I_1 passed through test 10; I_1 and III_0 so that I_2 passed through 14. No further rotations were made. Tests 1, 2, and 9 are left with small negative loadings in factor II in order not to reduce beyond reason the loadings of factor I in tests 14 and 15. The loading of factor III in test 9 before rotation, viz., 0.257, was found significant at $P=0.02$ by Lawley's unpublished method.

With the kind consent of Dr. Lawley, and by courtesy of the Royal Society of Edinburgh, we are able to give here the formula for the variance of a factor loading. The variance of the r th factor loading in test i is

$$\frac{\theta_r}{N} \left\{ 1 - \sum_{s=1}^r \theta_s l_{is}^2 + \frac{1}{2} \theta_r l_{ir}^2 \right\},$$

where N is the number of cases in the sample, l_{is} is the loading in test i of the particular factor s , and $\theta_r = 1 + \frac{1}{k_r^2}$, k_r^2 being the quantity obtained in the computation of the factor loadings and denoted by $h_1, k_1, \dots$, in Lawley's article (8, p. 173).

For example, denoting factors by Roman numerals and tests by Arabic, we have for the variance of the third factor in test 8,

$$\sigma_{8III}^2 = \frac{\theta_{III}}{N} \left\{ 1 - \theta_I l_{8I}^2 - \theta_{II} l_{8II}^2 - \frac{1}{2} \theta_{III} l_{8III}^2 \right\}.$$

The formula applies to the loadings before rotation and only when the number of cases in the sample is large, say, 200 or more. It ignores errors in estimating the communalities. The effect of these errors is in general small, however, especially when the number of tests is moderately large.

For both sexes the first rotated factor may be taken as g , the second as v , and the third as k , the space factor. Exact parallelism between the space-factor loadings for boys and girls is not to be expected, since the girls' loadings are on the borderline of significance and are thus garbled by random effects. However, the agreement in tests 8, 11, 6, 4, and 10 (in the order in which they are given after rotation) is quite satisfactory.

The g loadings of tests 14 and 15 are small, and could not be increased by changing the positions of axes without introducing negative loadings in other variables.

If the space factor is associated with the power of visual imagery, it is easy to understand its appearance in these picture tests, in particular as items in tests 5, 8, 9, and 10 involve a change of scale as between premiss and response. It is, however, surprising that the factor has been so readily detected as the tests comprise only 15 questions each with consequent low intercorrelations and communalities. That the spatial ability is more pronounced amongst boys than girls is doubtless due to an environment more favourable to its development amongst boys.

The sub-tests finally selected for Dr. Mellone's picture intelligence test, M.H.T. Pic. 1, were tests 2, 4, 5, 7, 8, 9, 10, 11, and 12 of the battery, that is, those tests with the highest communalities.¹ The correlation of scores in this test at 7+ with scores in a Moray House verbal intelligence test, M.H.T. 44, at 11+ was 0.681 for 7,028 children, boys and girls in approximately equal proportions. The standard deviations of the intelligence quotients in the two tests were 13.34 and 11.62 respectively. After correction to a standard deviation of 15 the estimated correlation coefficient for the population was 0.759. This correlation is not high, for an 11+ verbal intelligence test has a correlation with a similar 13+ test given two years later of 0.90 to 0.93, and with an interval of four years a correlation of at least 0.85 might be expected. The result thus supports the conclusion that the picture test involves a factor not shared by the verbal test. Whether this factor is the same as the group factor found in the Moray House space test is the subject of an enquiry now in progress.

¹ The test together with instructions and norms is available from the University of London Press.

VII. DREW'S 1947 ENQUIRY

L. J. Drew (2) found evidence of a space factor in boys at the ages of 11, 12, 13, and 16. His statistical method was, however, adversely criticized by P. Slater (16). It was accordingly thought well to make our own analysis of Drew's 11+ correlations by Lawley's method. First approximations for the first two factors were obtained by a centroid analysis, and our figures were in exact agreement with Drew's. The Lawley analysis yielded three significant factors only, not four as found by Drew; the second residual correlations gave $P < 0.001$, while the third gave $P = 0.30$. The results of the analysis are shown in Table VI.

TABLE VI. FACTOR ANALYSIS OF DREW'S BATTERY

181 boys; mean age, 11:9; mean I.Q., 100.37; I.Q. range, 90 to 116; first-year pupils in a 'senior (elementary) school.'

Tests	Before rotation				After rotation		
	I ₀	II ₀	III ₀	h ²	I ₂	II ₁	III ₁
Verbal Intell., M.H.T. .. 1	.514	.455	.314	.570	.402	.639	.000
Verbal test, Spearman .. 2	.381	.545	.035	.444	.084	.659	.050
Spatial test, Spearman .. 3	.757	.016	-.361	.704	.133	.359	.746
Perceptual test, Spearman 4	.649	.244	-.062	.485	.242	.513	.404
Alexander's { Passalong .. 5	.506	-.198	.433	.484	.676	.054	.154
{ Koh's Blocks 6	.782	-.400	.035	.774	.575	.000	.666
{ Cubes .. 7	.463	-.243	.243	.333	.516	-.006	.258
Teachers' Verbal Rating .. 8	.488	.345	.068	.362	.226	.529	.174
Teachers' Practical Rating 9	.486	.146	-.259	.324	.026	.351	.448
Variance, $S(I^2)$	2.960	.962	.553	4.475	1.351	1.641	1.487
Per cent. of total variance	32.89	10.69	6.14	49.72	15.01	18.23	16.52

The loadings before rotation agree reasonably well with Drew's loadings, while the polarity of the tests in respect of the second and third factors is the same for both analyses. When axes were rotated there was little choice of position if negative loadings were to be eliminated and axes were to remain orthogonal. Axes I₀ and II₀ were rotated so that I₁ passed through test 6; and I₁ and III₀ so that I₂ passed through test 1. Some of the results are unexpected; for example, Spearman's verbal test, test 2, appears as having practically no g , and Alexander's Passalong test as almost free from the third factor, which we take as k . Drew's rotation gives widely different results since he assumes that Spearman's perceptual test, test 4, is a true measure of g , whereas we have found in this and other work that it is a composite of three factors.

Drew's 11+ tests were apparently not well suited to his subjects, or were indifferent measures of the abilities in question, for the mean of his 36 correlation coefficients was only 0.304, and the communalities of the tests were correspondingly low. In spite of this the third factor, unrotated, is most pronounced, its loading in test 3, Spearman's spatial test, being 6.7 times its standard error. There is no evidence that a fourth factor such as Alexander's F is required to explain the correlations.

VIII. EL KOUSSY'S 1935 ENQUIRY

In his 1935 monograph A. H. El Koussy (5) brought forward weighty evidence in support of a space factor amongst boys aged 11 to 13 years. It is surprising that his work has not received more attention. At the time but crude methods of factor analysis were available, and his analysis involved suppositions that certain tests measured one and only one of the common factors. He found that eight tests of his battery of 28 had a group factor, which he called k .

We have recently analysed Koussy's correlations by Thurstone's centroid method, and have found three significant factors by applying McNemar's rough test of significance to the second residual correlations. After rotation factor loadings exceeding 0.4 in a group factor presumed to be k were found in 17 tests of the battery.

We do not present details of the analysis as the test of significance is open to doubt. It is proposed to reduce the number of variables to manageable proportions, from 28 to perhaps 12, by pooling tests which have an apparently similar factorial composition, and then to apply Lawley's method of analysis followed by his test of significance.

IX. SLATER'S 1943 ENQUIRY

Patrick Slater and Elizabeth Bennett (15) found no evidence of a space factor in spatial tests at either 11+ or 13+. Their work has received considerable attention on account of its bearing on the selection of children for technical courses. In view of its inconsistency with Drew's and Koussy's findings, the writer analysed Slater's 11+ correlations by Lawley's method. A significant third factor associated with the spatial variables was found.

Slater's correlations are based on scores from 211 boys and girls, aged 11+. The inclusion of girls doubtless lessened the chance of finding a significant space factor. We reduced his 17 variables to nine in number by pooling the correlations of similar tests in order to ease the computations, which otherwise would have been prohibitive. Weighting of the pooled tests was not possible as Slater records no standard deviations; each test was therefore given the same weight. We ended up with three non-verbal, three verbal, and three spatial variables as indicated in Table VII. The only test in the battery involving three-dimensional space perception was the spatial test, Part 4, numbered 11 on p. 144 of Slater's article and, rather irritatingly, 12 on p. 147. Such a test might be expected to have a high space-factor loading; but as its communality was only 0.128 as found by Slater, it can have discriminated but slightly amongst 11+ children. For this reason it was not treated as a variable by itself, and was pooled with two other space tests.

It may be doubted whether the tests as a whole were suited to the children tested. In the absence of means and standard deviations of raw scores one can only judge from the correlations, which are low. The intercorrelations of the 17 tests at 11+ averaged only 0.345; the same 17 tests at 13+ gave an average intercorrelation of 0.479. Thus either the tests were poorly adapted to the 11+ children or the abilities tested were more specific at the earlier age, and the latter supposition is not likely to gain much support. Table VII gives the pooled correlations and the resulting factorial matrix.

TABLE VII. CORRELATIONS AND FACTOR LOADINGS OF SLATER'S VARIABLES AFTER POOLING

A. CORRELATIONS

Test No.*	Tests	1	2	3	4	5	6	7	8	9
1	Non-verb., N.I.I.P. 70, Pt. 1 .. 1	—	·523	·395	·471	·346	·426	·576	·434	·639
2, 3	„ Pts. 2 and 3 .. 2	·523	—	·479	·506	·418	·462	·547	·283	·644
4	„ Pt. 4 .. 3	·395	·479	—	·355	·270	·254	·452	·218	·504
12, 13	Verbal, Slater, Pts. 1 and 2 .. 4	·471	·506	·355	—	·691	·791	·443	·285	·505
14, 15	„ Pts. 3 and 4 .. 5	·346	·418	·270	·691	—	·679	·382	·149	·409
16, 17	„ Pts. 5 and 6 .. 6	·426	·462	·254	·791	·679	—	·372	·314	·472
5, 6	Spatial, Squares and Designs .. 7	·576	·547	·452	·443	·382	·372	—	·385	·680
7, 8	„ Pt. 1 and Shapes .. 8	·434	·283	·218	·285	·149	·314	·385	—	·470
9, 10, 11	„ Pts. 2, 3, and 4 .. 9	·639	·644	·504	·505	·409	·472	·680	·470	—

* This numbering refers to that on p. 144 of Slater's article. A different numbering is used on p. 147.

B. FACTOR LOADINGS

Tests		Before rotation				After rotation		
		I ₀	II ₀	III ₀	h ²	I ₂	II ₁	III ₁
Non-verbal, Pt. 1 .. 1	.. 1	·668	·313	·111	·557	·639	·071	·379
„ Pts. 2 and 3 .. 2	.. 2	·693	·241	—·172	·568	·731	·145	·110
„ Pt. 4 .. 3	.. 3	·501	·296	—·236†	·394	·628	·000	·000
Verbal, Pts. 1 and 2 .. 4	.. 4	·840	—·314	—·028	·805	·532	·698	·186
„ Pts. 3 and 4 .. 5	.. 5	·704	—·322	—·143	·619	·463	·635	·033
„ Pts. 5 and 6 .. 6	.. 6	·804	—·372	·107	·796	·426	·729	·288
Spatial, Squares and Designs .. 7	.. 7	·669	·390	—·058	·602	·739	·005	·237
„ Pt. 1 and Shapes .. 8	.. 8	·448	·266	·409†	·439	·329	—·001	·575
„ Pts. 2, 3, and 4 .. 9	.. 9	·770	·421	·019	·771	·806	·030	·347
Variance, $S(l^2)$	..	4·266	·984	·302	5·552	3·322	1·450	·781
Per cent. of total variance ..	..	47·41	10·93	3·35	61·69	36·91	16·11	8·68

† The loading of III₀ in Test 8 is 4·79 times its standard error and that in Test 3 2·35 times its standard error.

The correlations were factorized by Lawley's maximum likelihood method. A third factor was found, significant at $P=0.05$. Axes I₀ and II₀ were rotated so that I₁ passed through test 3; and then axes I₁ and III₀ so that I₂ passed through test 3.

The first factor after rotation may be taken as g , the second factor is obviously v , and the third factor no less obviously k . Test 1, although designed as a non-verbal intelligence test, involves the matching of shapes and so the presence of an appreciable k loading need cause no surprise. Lawley's significance test of the unrotated third factor loadings in tests 3 and 8 shows a high degree of significance for each test, though the loadings are of opposite sign.

Slater gives in the article in question a comparison of the factor loadings at 11+ and at 13+ in the same tests. If, as we have reason to believe, the tests were less efficient measures of the abilities at 11+, the results of the comparison can have little weight.

X. SUMMARY

1. Boys have been shown to be significantly superior to girls in a variety of space and performance tests, the same girls often showing a superiority in verbal tests. These sex differences provide presumptive evidence of a special spatial ability.

2. A battery of nine tests, including a Moray House space test, has been subjected to factorial analysis by D. N. Lawley's method of maximum likelihood. Four significant factors were found, of which three were evaluated. After rotation of axes the loadings of a third factor in the space test was of the same order of magnitude as that of the verbal factor usually found in verbal intelligence tests. A higher loading in items with three-dimensional objects was found than in items with two-dimensional objects.

3. A factorial analysis by Lawley's method of the sub-tests of a 7+ picture intelligence test by Margaret Mellone indicates a significant group factor other than verbal amongst both boys and girls.

4. The correlations obtained by L. J. Drew with 11+ boys from nine variables have been analysed by Lawley's method. Three significant factors were found, but not four. After rotation considerable loadings in a factor confined to Spearman's spatial and perceptual tests and Alexander's performance battery appeared.

5. El Koussy's battery of 28 tests has been analysed by the centroid method. Three significant factors were found.

6. The correlations obtained by Patrick Slater in a mixed group of 11+ boys and girls have been analysed by Lawley's method. A third significant factor, not found by Slater, has been established ($P=0.050$), one spatial test having a highly significant loading ($P<0.0001$).

7. It now remains to see whether later performance in 'technical' subjects is successfully predicted by a space test: that is, whether the test in a mixed battery of other predicting tests has a significant partial regression coefficient. Enquiries to this end are already in train.

ACKNOWLEDGEMENTS

First of all the writer is greatly indebted to Professor Godfrey Thomson who presided over the Moray House enquiry from beginning to end and supplied refreshing stimuli at times of quiescence caused by the pressure of routine work. Next very much is owed to Dr. D. N. Lawley for his invaluable and ever-ready help in untying many knotty statistical points. And finally, the writer and his colleagues gratefully acknowledge grants from the Godfrey Thomson Research Fund which defrayed the expenses of the enquiry.

REFERENCES

1. Bain, J. T. (1946). *The Construction of a Space Test*. B.Ed. Thesis. Moray House, University of Edinburgh.
2. Drew, L. J. (1947). The measurement of technical ability. *Occ. Psych.*, XXI, 34.
3. Holliday, F. (1940). The selection of apprentices for the engineering industry. *Occ. Psych.*, XIV, 69.

4. Holliday, F. (1941). A further investigation into the selection of apprentices for the engineering industry. *Occ. Psych.*, XV, 173.
5. Koussy, A. H. El (1935). The visual perception of space. *Brit. J. Psych. Mon. Suppl.* XX.
6. Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edin.*, LX, 64.
7. Lawley, D. N. (1942). Further investigations in factor estimation. *Proc. Roy. Soc. Edin.* (Section A), LXI, 176.
8. Lawley, D. N. (1943). The application of the maximum likelihood method to factor analysis. *Brit. J. Psych.*, XXXIII, 172.
9. MacMeeken, A. M. (1939). *The Intelligence of a Representative Group of Scottish Children*. University of London Press.
10. Mellone, Margaret A. (1943). *A Factorial Study of Picture Tests for Young Children*. Ph.D. Thesis. Moray House, University of Edinburgh.
Also (1944) *Brit. J. Psych.*, XXXV (1), 9.
11. Mills, L. F. (1947). *The Properties of a Space Test*. B.Ed. Thesis. Moray House, University of Edinburgh.
12. Scottish Council for Research in Education (1933). *The Intelligence of Scottish Children*. University of London Press.
13. Shuttleworth, C. W. (1942). Tests of technical aptitude. *Occ. Psych.*, XVI, 175.
14. Slater, Patrick (1941). Tests for selecting secondary and technical schoolchildren. *Occ. Psych.*, XV (1), 10.
15. Slater, Patrick, and Bennett, Elizabeth (1943). The development of spatial judgment and its relation to some educational problems. *Occ. Psych.*, XVII (3), 139.
16. Slater, Patrick (1947). Evidence on selection for technical schools. *Occ. Psych.*, XXI, 135.
17. Thomson, Godfrey H. (1946). *The Factorial Analysis of Human Ability*. University of London Press.
18. Burt, Cyril (1943). The education of the young adolescent. *Brit. J. Educ. Psych.*, XIII, 132.

MULTIVARIATE ANALYSIS APPLIED TO DIFFERENCES BETWEEN NEUROTIC GROUPS

By C. RADHAKRISHNA RAO and PATRICK SLATER

National Foundation for Educational Research

I. Introduction. II. Source of the Material and Definition of the Problems. III. Analysis of Dispersion. IV. The Generalized Differences between the Groups. V. Evidence of Variation between the Groups in more than One Dimension. VI. Statistical Criterion to determine the Group to which an Individual belongs. VII. Summary.

I. INTRODUCTION

Psychologists are accustomed to representing variations in abilities and personality traits as quantitative measurements in different dimensions. This device is used in the multiple correlation procedures often applied to problems of vocational selection, in factor analysis, and so on. It seems more reasonable to extend the same treatment to personality traits which aid in distinguishing neurotic individuals from normal, than to make any special exception of them. This approach was adopted by Sir Cyril Burt in a factorial analysis of temperamental variations observed among a group of delinquent and neurotic children (3), and has since been widely used for problems in the psychiatric field, for instance, by Eysenck and his colleagues (4).

For a comprehensive presentation of the grounds for applying quantitative methods to problems of psychiatric diagnosis we may make particular reference to a paper by Slater and Slater (14), collating evidence from the fields of psychology, psychiatry, and genetics and showing its convergence upon this point. The characteristics which differentiate neurotics from normals may, in their view, be treated as continuous variables. In each of the dimensions considered the differences between neurotics and normals and between different groups of neurotics would, they hold, be found to be differences solely in degree. This would account for many of the observed phenomena, e.g., that the incidence of neuroses in different military duties varies directly with the amount of hazard encountered, that neurotic groups fade off into one another clinically and do not form qualitatively distinct groups, that among the relatives of patients suffering from one form of neurosis the incidence of other forms of neurosis is exceptionally high, etc. According to this view the apparent qualitative distinction between neurotics and normals is the resultant of a balance between two quantitatively variable forces. When stress exceeds resistance, the man becomes classifiable as neurotic, when resistance exceeds stress, as normal. Both stress and resistance vary in degree and in kind.

In the present paper we make use of this quantitative approach ; and in view of the existing evidence on its reasonableness and usefulness, we feel that we do not need to argue its justification again. Our data consist of three kinds of psychological measurements applied to 256 men who have been classified into five neurotic groups and a control group of normals. We discuss the extent to which the groups can be distinguished in terms of the measurements, the main dimensions of variation found between the groups, and the appropriate procedure for sorting men whose measurements are known into the groups to which they are most likely to belong.

The methods we have used are general and may prove to have other useful psychological applications, for instance, in selection. In developing a selection procedure for an occupation, it is usual to apply psychological tests to a group of entrants ; and after allowing them to remain in the occupation long enough to provide a reasonably reliable indication of the degree of success each has

attained, to correlate their scores on the tests with the criterion of success and calculate a multiple regression equation. This defines a method of 'weighted summation' to apply to individuals' scores; and it is generally assumed that the higher the summed score obtained by any individual, the greater is his probability of succeeding in the occupation. This assumption (unless supplemented by others) overlooks the fact that people are often unsuccessful in occupations too far beneath their abilities. In considering an individual's suitability it might be better to examine the multivariate dispersion of the test performances of the men who prove to be normally successful in the occupation. Their averages will indicate a point in the multidimensional space defined by the test scores, and the dispersion will indicate a region of permissible variation about that point. Then the likelihood that an individual will succeed in the occupation, i.e., form a homogeneous member of the group of people who are normally successful in it, can be determined by relating the point in the multidimensional space indicated by his test scores to the region of permissible variation about the multivariate mean point of the normally successful group.

In view of these possible extensions, we have confined our discussion of the psychological interpretation of our data and our results to a broad outline, and have concentrated our attention on exhibiting the procedure used as clearly as possible, quoting references for proofs and for working methods not fully explained. A fuller psychological discussion of the results is presented in (10).

II. SOURCE OF THE MATERIAL AND DEFINITION OF THE PROBLEMS

Data concerning 256 officers from the Army and the Navy were collected by W. Mayer Gross and J. N. P. Moore; 55 were Army officers on active service, whose records provide a control group of normal cases. The remaining 201 were referred to hospital for the treatment of neuroses, 100 from the Army and 101 from the Navy. The diagnoses of these cases are shown in Table I.

Mayer Gross and Moore noted the presence or absence of certain pointers in each case. An analysis of these observations was made by Patrick Slater (15). Thirteen of the 'pointers' were found to occur significantly more frequently among the neurotic officers than among the normal, viz.: 1. Hereditary predisposition; 2. Physical ill health; 3. Neurotic traits in childhood; 4. Former psychiatric illnesses; 5. Shy, markedly introverted in childhood or adolescence; 6. Difficulty in making social contact; 7. Emotional instability; 8. Obsessional features; 9. Apprehensiveness; 10. Dependence in childhood and later life; 11. Unstable work record; 12. Marriage and sex difficulties; 13. Alcoholism.

The precise definition of each 'pointer,' used in deciding whether it should be recorded as present or absent in each case, is reported separately (10).

Factor analysis was applied to the frequencies with which the pointers occurred concomitantly (15) and disclosed the presence of two clusters of particularly closely associated pointers within the general group. Pointers 4, 7, and 13 form the first of these clusters; the most likely explanation is that they are connected as evidences of instability. Prolonged psychiatric illnesses and chronic alcoholism are not to be expected in the histories of men holding commissions in the Army and the Navy. Where these pointers are noted they must usually indicate disturbances of an episodic character. The other cluster includes only pointers 5 and 6, which are pronouncedly similar in definition, although adolescent shyness is sometimes overcome in adult life.

The simplest method of applying the information from all the pointers is to take the total number found in each case as a score. On the other hand the greatest discriminatory value will be obtained when optimum weights are calculated for each pointer separately. Results of intermediate value will be obtained by dividing them into a few classes and assigning a separate weight to each class. The classification indicated by the factor analysis was a possible one to adopt, and an experiment seemed worth making to discover how far the discriminatory value of the pointers could be enhanced by adopting it. The pointers were therefore divided as follows: *A.* Nos. 1, 2, 3, 8, 9, 10, 11, and 12; *B.* Nos. 4, 7, and 13; *C.* Nos. 5 and 6. The number of pointers of each class noted in each case was used as a score, giving three scores for each individual. The first, *A*, may be described as a measure of constitutional inadequacy, the second of instability, and the third of shyness.

The total and the average scores of the groups are shown in Table I.

TABLE I. TOTAL AND AVERAGE SCORES

Group	Sample size (<i>n</i>)	Scores					
		<i>A</i>		<i>B</i>		<i>C</i>	
		Total	Mean	Total	Mean	Total	Mean
Anxiety state	114	334	2.9298	133	1.1667	83	.7281
Hysteria	33	100	3.0303	41	1.2424	18	.5455
Psychopathy	32	122	3.8125	59	1.8438	26	.8125
Obsession	17	80	4.7059	27	1.5882	19	1.1176
Personality change ..	5	7	1.4000	1	.2000	0	0.0000
Normal	55	33	.6000	8	.1455	12	.2182
Total	256	676		269		158	

The most conspicuous single difference is between the normal cases and the neurotics, but there are also marked differences between the groups of neurotics. The five cases of post-traumatic personality change approximate to the normal cases. Closely similar to one another, but distinctly different from the normal and the remaining neurotic groups are the hysterics and anxiety states. Still further removed from normal are the obsessional and psychopathic cases; but between these two groups some differences appear to exist. Whereas the obsessionals exhibit more pointers per person than the psychopaths, this difference is wholly due to an excess of symptoms of inadequacy and shyness; there is less evidence of instability among them than among the psychopaths. This suggests that in addition to variations in the degree to which the neurotic groups differ from normal, a further source of variation may be found and may prove useful for differential diagnosis.

Two problems thus arise:

1. Is there sufficient evidence to demonstrate variation between the groups in more than one dimension, or can the observed differences between them be treated as differences simply in degree?

2. What is the best method of assigning an individual to one of these groups given his scores in *A*, *B*, and *C*?

III. ANALYSIS OF DISPERSION

The first step in problems of this nature is to test whether the observed differences between the groups are significant. If only a single score is available the differences between the groups can be tested by the method of analysis of variance. In this case we analyse the total sum of squares into 'between' and 'within' groups and compare the mean squares derived from them. In the case of multiple scores both the total sums of squares and products can be analysed into 'between' and 'within' groups and the problem is that of comparing simultaneously the mean squares and products derived from them (1, 12, and 16).*

* This method has been termed the analysis of dispersion to distinguish it from the variance-covariance analysis used in the elimination of concomitant variation. A further discussion is given in the above-mentioned paper (12), wherein this concept has been used in deriving unbiased tests of significance in multivariate analysis.

The total sum of products matrix (designated for brevity an S.P. matrix) is given in Table II.

TABLE II. THE TOTAL PRODUCT-SUM MATRIX WITH 255 D.F.

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	942.9375	224.6719	194.7812
<i>B</i>	..	224.6719	228.3398	41.9766
<i>C</i>	..	194.7812	41.9766	166.4844

The terms in this matrix are the sums of squares and products $\sum_i^2 - (\sum_i)^2/N$, and $\sum_{ij} - (\sum_i)(\sum_j)/N$. N being the total number of cases, over which the summation is made without regard to grouping, and i and j being used to denote the measurements in each variable in turn. In this case $N = 256$, so the degrees of freedom for the matrix in Table II are $N - 1 = 255$. The total matrix can be divided 'between' and 'within' groups, as in Tables III and IV.

TABLE III. THE PRODUCT-SUM MATRIX BETWEEN GROUPS WITH 5 D.F.

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	367.7248	161.7773	76.2389
<i>B</i>	..	161.7773	76.4732	33.0330
<i>C</i>	..	76.2389	33.0330	17.7109

The terms in this table are computed from the data in Table II as $\frac{334^2}{114} + \frac{100^2}{33} + \frac{122^2}{32} + \frac{80^2}{17} + \frac{7^2}{5} + \frac{33^2}{55} - \frac{676^2}{256}$ for variable A , $\frac{133 \times 334}{114} + \frac{41 \times 100}{33} + \frac{59 \times 122}{32} + \frac{27 \times 80}{17} + \frac{1 \times 7}{5} + \frac{8 \times 33}{55} - \frac{269 \times 676}{256}$ for variables A and B , and similarly for others. They have 5 d.f.—one less than the number of groups.

TABLE IV. THE PRODUCT-SUM MATRIX WITHIN GROUPS WITH 250 D.F.

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	575.2127	62.8946	118.5423
<i>B</i>	..	62.8946	151.8666	8.9436
<i>C</i>	..	118.5423	8.9436	148.7735

The terms in this table are the differences between the two corresponding terms in Tables II and III. They have 250 d.f., one degree of freedom being sacrificed from the number of cases in each group. To test for overall differences between the groups the following ratio is calculated:

$$\Lambda = \frac{\text{determinant of S.P. matrix 'within' groups (Table IV)}}{\text{determinant of S.P. matrix for the total (Table II)}} = \frac{10360977}{20791385} = .49833$$

We may use Bartlett's approximation (1) to calculate $\chi^2 = -\{n - \frac{1}{2}(m + q + 1)\} \log_e \Lambda$, where n is the d.f. of the denominator, q that of the numerator, and m is the number of variables under consideration. Here $\chi^2 = [255 - \frac{1}{2}(3 + 5 + 1)] \times .696493 = 175.1680$. Entering this value in a χ^2 table with $m \times q = 15$ degrees of freedom, we find that the differences are significant.

Having established that all groups are not the same, we need to ascertain whether the main difference between the normals and personality changes on the one hand and the remaining neurotic groups on the other is the only one which is statistically significant, or whether differences among the remaining neurotic groups are too great to be overlooked. Since the significance of the former cannot be doubted we need only test for differences in the four other neurotic groups. The S.P. matrix with 3 d.f. for them is given in Table V.

TABLE V. THE PRODUCT-SUM MATRIX BETWEEN THE FOUR NEUROTIC GROUPS :
Anxiety States, Hysteries, Psychopaths, and Obsessions

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	59.4322	22.2319	12.5026
<i>B</i>	..	22.2319	12.8718	3.6374
<i>C</i>	..	12.5026	3.6374	3.8532

The S.P. matrix for within all the groups + between the above four groups is given in Table VI.

TABLE VI. THE PRODUCT-SUM MATRIX WITHIN ALL GROUPS + BETWEEN FOUR NEUROTIC GROUPS

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	634.6449	85.1265	131.0449
<i>B</i>	..	85.1265	164.7384	12.5810
<i>C</i>	..	131.0449	12.5810	152.6267

This has $250 + 3 = 253$ d.f. The ratio is

$$\Lambda = \frac{\text{determinant of matrix in Table IV}}{\text{determinant of matrix in Table VI}} = \frac{10360969}{12202393} = .84909; \log_e \Lambda = -.16359$$

$$\chi^2 = [253 - \frac{1}{2}(3 + 3 + 1)] \times 0.16359 = 40.82$$

The value as a χ^2 with $3 \times 3 = 9$ d.f. is significant at the 1 per cent. level thus establishing significant differences among the four neurotic groups as well. The situation demands a closer study of the differences which is attempted below.

IV. THE GENERALIZED DISTANCES BETWEEN THE GROUPS

The three variables *A*, *B*, and *C* can be considered as defining a space of three dimensions so that any individual with assigned scores can be represented by a point in such a space. The individuals belonging to any group will then be represented by a cluster of points round the point representing the mean values of the scores. If the mean values of two groups are close together then the clusters of points corresponding to them will overlap to some extent. Rao (11) has shown that the degree of overlap between any two such clusters can be measured by the quantity *

$$D^2 = \sum_{i=1}^{i=3} \sum_{j=1}^{j=3} a^{ij} d_i d_j,$$

* Professor Burt points out that this is essentially the same as Mahalanobis' 'generalized distance' (Mahalanobis, P. C., 'On the generalized distance in statistics' *Proc. Nat. Sec. Ind.*, XII, 1936, pp. 493).

where d_1, d_2, d_3 are the differences between the mean values of the measurements A, B , and C respectively in the two groups and a^{ij} are the elements of the matrix reciprocal to the within group dispersion matrix (a_{ij}). The higher the value of D , the smaller will be the amount of overlap so that this measure may be regarded as the distance between two groups. The computational procedure for evaluating these distances is given below.

In our example the S.P. matrix within groups of Table IV has 250 d.f. Dividing each element of this matrix by 250, we have the dispersion matrix within groups as given in Table VII. It might be noted that we are assuming the equality of the dispersion matrices in each group and estimating its elements. A different technique might be needed if the dispersion matrices in the various groups differ widely. The reciprocal of this matrix is shown in Table VIII.

TABLE VII. THE DISPERSION MATRIX WITHIN GROUPS (a_{ij})

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	2.300851	.251578	.474169
<i>B</i>	..	.251578	.607466	.035774
<i>C</i>	..	.474169	.035774	.595094

TABLE VIII. THE RECIPROCAL OF THE DISPERSION MATRIX (a^{ij})

		<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	..	.543234	-.200195	-.420813
<i>B</i>	..	-.200195	1.725807	.055767
<i>C</i>	..	-.420813	.055767	2.012357

To calculate the D^2 between the groups (say) normal and personality changes we find the differences d_1, d_2, d_3 in mean values of A, B, C from Table I, viz., $d_1 = .8000, d_2 = .0545, d_3 = -.2182$. Using the elements (with four significant figures) of the matrix in Table VIII, D^2 is evaluated as $D^2 = .5432 \times (.8000)^2 + 1.7258(.0545)^2 + 2.0123(-.2182)^2 + 2(-.2002)(.8000)(.0545) + 2(-.4208)(.8000)(-.2182) + 2(.0558)(.0545)(-.2182) = .5767$. Table IX gives all possible values of D^2 and enables each group to be compared with every other. With respect to each group in turn all the others are arranged in increasing order of D^2 .

TABLE IX. THE VALUES OF D^2 ARRANGED IN INCREASING ORDER OF MAGNITUDE WITH RESPECT TO EACH GROUP †

Normal (N)		Personality Change (P.C.)		Anxiety State (A.S.)		Hysteria (H)		Psychopathy (P)		Obsession (O)	
Group	D^2	Group	D^2	Group	D^2	Group	D^2	Group	D^2	Group	D^2
PC	.5767	N	.5767	H	.0933*	AS	.0933*	O	.5870*	P	.5870*
AS	3.3773	AS	2.4999	P	.9332	P	.7538	H	.7538	H	1.3735
H	3.8038	H	2.5524	O	1.4619	O	1.3735	AS	.9332	AS	1.4619
P	7.6159	P	6.0649	PC	2.4999	PC	2.5524	PC	6.0649	PC	7.0023
O	9.0430	O	7.0023	N	3.3773	N	3.8038	N	7.6159	N	9.0430

* No significant difference.

† The following symbols have been employed to represent the various groups: Normal = N, Personality change = PC, Anxiety state = AS, Hysteria = H, Psychopathy = P, Obsession = O.

The values of D^2 as given in Table X are the estimates of the square of the distances in the populations. To calculate the chance of a D^2 exceeding the observed when two groups are the same we can use $F = \frac{n_1 n_2}{n_1 + n_2} \cdot \frac{1}{n} \cdot \frac{n - m + 1}{m} D^2$, as the variance ratio with m and $n - m + 1$ d.f. Here m is the number of variates, n the number of d.f. of the estimates in the error dispersion matrix, n_1 and n_2 the numbers in the two samples. This test is due to Hotelling (9). It is identical with the test of significance of the discriminant function given by Fisher (6). In our case $m = 3$ and $n = 250$. The values of F are given in Table X.

TABLE X. THE VARIANCE RATIOS FOR ALL PAIRS OF GROUPS

		N	PC	AS	H	P	O
N ..	..	—	.87*	41.43	25.94	50.95	28.83
PC ..	..		—	3.96	3.66	8.67	8.95
AS ..	..			—	.79*	7.71	7.15
H ..	..				—	4.05	5.09
P ..	..					—	2.15*
O ..	..						—

* Indicates ratios not significant at the 5 per cent. level.

We find all the values of D^2 are significant at the 5 per cent. level except those between normal and personality change, anxiety and hysteria, psychopathy and obsession. It might be noted that non-significance does not imply that the two groups involved are identical. They are likely to be differentiated when further evidence is accumulated. Another important observation that can be made from Table X is that the ratio for obsession and psychopathy is greater than that for anxiety state and hysteria, although the numbers of cases represented in the latter two groups are greater than those for the former two. This indicates that there is greater difference between obsession and psychopathy than between anxiety state and hysteria. Samples of about 60 each for obsession and psychopathy may reveal their differences, but a larger size would be needed for anxiety state and hysteria.

Table IX shows that the six groups can be arranged approximately in a linear order with normal and obsessional groups at the extremes and personality change, anxiety state, hysteria, and psychopathy coming in between in the above order. The pairs normal and personality change, anxiety state and hysteria, psychopathy and obsession are closer to one another in this linear arrangement.

V. EVIDENCE OF VARIATION BETWEEN THE GROUPS IN MORE THAN ONE DIMENSION

The evidence given so far does not demonstrate that the variation between the groups lies in more than one dimension. It is consistent with the possibility that the points indicating the mean positions of the six groups do not diverge significantly from the nearest corresponding points on one straight line passing through the three dimensional space defined by the three variables. If this is so, the evidence of these measurements would be insufficient to show that the neurotic states differ from one another in any way except in degree. Predisposing conditions would appear to give

rise to neuroses of different kinds as they become more severe in their total effect, but they would not appear likely to produce different neuroses when they occur in different combinations. Is this so, or do different combinations produce different effects? Does the variation extend significantly in more than one direction?

To consider the hypothesis that the differences between the groups can be attributed to a single linear function, let k_1, k_2, k_3 denote the direction cosines of the line. Fisher (8) has described a method for calculating the values of k_1, k_2, k_3 which define the line giving the closest fit to the mean values of the groups. This involves solving the determinantal equation for φ^* in Table XI.

TABLE XI. THE DETERMINANTAL EQUATION FOR φ

$$\begin{vmatrix} 367.7248 - \varphi & 575.2127 & 161.7773 - \varphi & 62.8946 & 76.2389 - \varphi & 118.5423 \\ 161.7773 - \varphi & 62.8946 & 76.4732 - \varphi & 151.8666 & 33.0330 - \varphi & 8.9436 \\ 76.2389 - \varphi & 118.5423 & 33.0330 - \varphi & 8.9436 & 17.7109 - \varphi & 148.7735 \end{vmatrix} = 0$$

The matrix of the determinant in Table XI is the matrix of Table III minus φ times the matrix of Table IV. To simplify the computation of φ we may multiply the matrix of the determinant in Table XI by the matrix in Table VIII. This is equivalent to premultiplying the matrix of Table III by that of Table VIII and subtracting the matrix

$$\begin{matrix} 250\varphi & 0 & 0 \\ 0 & 250\varphi & 0 \\ 0 & 0 & 250\varphi \end{matrix}$$

for, as is shown by the computations already discussed, the matrix in Table VIII is 250 times the reciprocal of the matrix in Table IV. The product of the matrices is thus

TABLE XII. PRODUCT MATRIX

$$\begin{matrix} 135.2909 - \mu & 58.6725 & 27.3495 \\ 209.8339 & 101.4343 - \mu & 42.7341 \\ 7.7001 & 2.6617 & 5.4009 - \mu \end{matrix}$$

Here $\mu = 250\varphi$. The values of μ , and thus of φ , for which the determinant of this matrix is zero, are found by expanding the determinant. The equation for μ is thus found to be: $\mu^3 - 242.1261\mu^2 + 2365.848134\mu - 5455.7616654 = 0$. The three roots of this equation correct to four decimal places are: $\mu_1 = 232.0312$, $\mu_2 = 6.4481$, $\mu_3 = 3.6468$; total 242.1261.

The variations absorbed by these roots are respectively 95.8 per cent., 2.7 per cent. and 1.5 per cent.; so the second and third appear relatively unimportant.

The total variation, 242.1261, has $m(s-1) = 15$ degrees of freedom. Here m is the number of variables and s the number of groups compared. Since the degrees of freedom of the within group dispersion matrix are large the overall differences could be tested by using the total variation 242.1261 as a χ^2 with 15 degrees of freedom. This test is alternative to the Λ test employed in section 3 and both are equivalent in large samples. The observed value of χ^2 is significant at the 1 per cent. level.

We now ask the question whether the mean values of the six groups are collinear in the three-dimensional character space. If this is so the whole variation should be concentrated along the line of the best fit. A significant value of the residual variation would disprove the hypothesis of collinearity. The total variation is split into two groups, $242.1261 = 232.0312 + 10.0949$, corresponding to the first and the other roots, with the distribution of degrees of freedom: $m(s-1) = (m+s-2) + \dots$, i.e., $15 = 7 + 8$. The residual 10.0949 as χ^2 with 8 degrees of freedom is not significant; thus there is no evidence of variation in the other dimensions.

* The quantity φ and the square of the canonical correlation λ (2) are connected by the relation $\varphi = \lambda/(1 - \lambda)$.

If this χ^2 had been significant we could proceed to test for coplanarity and so on till a non-significant residual variation is reached. The distribution of the degrees of freedom among the various roots is $m(s-1) = (m+s-2) + (m+s-4) + \dots$, each term being two less than the previous one. The degrees of freedom of the residual in any case is the total minus the sum of degrees of freedom for the roots accounting for the total variation on a specified hypothesis.

Bartlett (1) has suggested another χ^2 approximation for testing the significance of the roots, based on a subdivision of the χ^2 calculated on p 20 (8), and illustrated its use in a recent paper (2). It consists in evaluating the following statistics :

Root	Term	χ^2	d.f.
First	$[255 - \frac{1}{2}(m+s)] \log_e(1 + \varphi_1) =$	165.14	7
Second	$[255 - \frac{1}{2}(m+s)] \log_e(1 + \varphi_2) =$	6.35	5
Third	$[255 - \frac{1}{2}(m+s)] \log_e(1 + \varphi_3) =$	3.62	3

The χ^2 for the second and third roots $= 6.35 + 3.62 = 9.97$, with $5 + 3 = 8$ degrees of freedom. It is not significant, and therefore indicates collinearity. While no exact test is known, in large samples the two tests outlined above are equivalent and either may be considered adequate.

In the present case only the first root is significant, but in problems of this nature it is of practical importance to consider at least some of the smaller roots in determining the configuration of the various groups. It might happen that the variation in the dimension corresponding to a smaller root is concentrated among a few of the groups, in which case very large samples would be necessary to establish significance. Any noticeable difference between two groups in the mean values of the canonical variate corresponding to such a root cannot be strictly interpreted as real when the overall test does not establish the significance of this root. But this additional analysis may throw some light on the prospects of future investigations. In the present case the canonical variates corresponding to the first two roots have been calculated and the configuration of the various groups is indicated in Fig. 1.

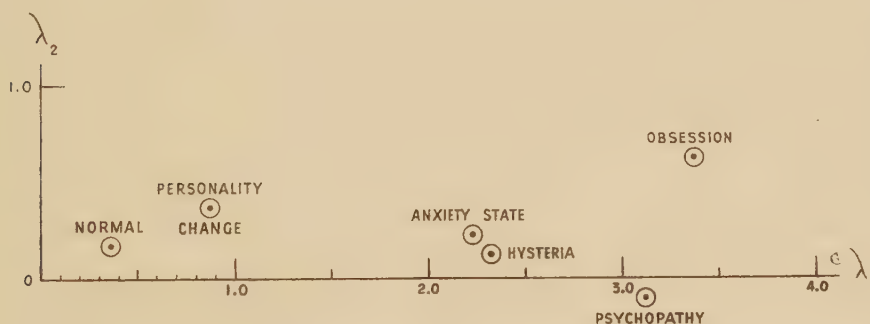


FIG. 1 The Configuration of the Several Groups.

The coefficients k_1, k_2, k_3 for the best linear fit are obtained from the equations

$$(135.2909 - 232.0312) k_1 + 58.6725 k_2 + 27.3495 k_3 = 0$$

$$209.8339 k_1 + (101.4343 - 232.0312) k_2 + 42.7341 k_3 = 0$$

$$7.7001 k_1 + 2.6617 k_2 + (5.4009 - 232.0312) k_3 = 0$$

The matrix of these equations is obtained from that of Table XII by substituting for μ the maximum root 232.0312. Putting $k_3 = 1$ arbitrarily we find that the proportional values of k_1 and k_2 are $k_1 = 18.84886$, $k_2 = 30.61225$.

The variance of $k_1A + k_2B + k_3C$ is $a_{11}k_1^2 + a_{22}k_2^2 + a_{33}k_3^2 + 2a_{12}k_1k_2 + 2a_{13}k_1k_3 + 2a_{23}k_2k_3$ where a_{ij} are the elements of the matrix in Table IV. Using the values of k_1, k_2, k_3 obtained as above we find the variance as 1697.6281 or the standard deviation as $\sqrt{1697.6281} = 41.2023$. Dividing k_1, k_2, k_3 by this value we find the standardized best linear functions $.4575A + .7430B + .0243C$.

Similarly the standardized linear function for the second dimension is $.4071A - 1.0473B + .3292C$. These functions are known as the first and second canonical variates. The mean values of these variates for all the groups are given in Table XIII.

TABLE XIII. MEAN VALUES OF THE FIRST TWO CANONICAL VARIATES

Group	λ_1	λ_2
Normal	.3879	.1637
Personality change	.7891	.3604
Anxiety state	2.2249	.2105
Hysteria	2.3227	.1120
Psychopathy	3.1339	-.1115
Obsession	3.3601	.6204

These values can be used for a pictorial representation of the groups as shown in Fig. 1. In the first dimension the normal group occupies a position at one end of the scale; the neurotic groups are spread out towards the other extreme, the small group of cases of post-traumatic personality change being the only one which is not clearly distinct from normal. At the other extreme are the psychopathic personalities and the obsessionals; in terms of this dimension they lie very close together. The anxiety states and the hysterias are also found to approximate closely to one another, but they lie a considerable distance from the other groups.

The preponderance of the part of the total variation between the groups which occurs in this dimension is the most striking finding in the analysis. The facts that it is the first of the dimensions found, and that each of the original scores makes a positive contribution to the variation in it, and, above all, that in it the normal group appears at one pole and the various neurotic groups diverge all towards the other—these facts suggest that it can be identified with the general factor among neurotic characteristics described by Eliot Slater (13), Eysenck (4), and other writers. Whether this general dimension of 'neuroticism' indicates the existence of any unitary psychological trait, or is simply a reflection of the fact that most neurotic characteristics are non-specific and found with varying degrees of frequency among all neurotic states, is a controversial question upon which it is unnecessary to enter here. Mayer Gross, Moore, and P. Slater prefer the latter alternative (10).

Although the variation in the second dimension is very much smaller, the arrangement of the groups it discloses invites some psychological consideration. The equation defining variation in this dimension contrasts the scores for inadequacy with the scores for instability by giving them opposite signs. At one extreme is the obsessional group, which in terms of average scores is the most highly inadequate but not the most unstable; at the other extreme is the psychopathic group—the most unstable but not the most inadequate. The psychological picture presented by this arrangement of the groups is a familiar one: obsessional cases are notoriously fixed in their habits; psychopaths notoriously irresponsible and unreliable. What is surprising is not that some contrast of this kind was found, but that the observations exhibit so little variation in this respect. But as the score is based on a summation of three pointers only, it is likely that variation in this dimension is insignificant than that it has been insufficiently accurately measured.

The D^2 (square of the generalized distance) between the obsessional and psychopathic groups due to the first and second canonical variates is .5766. The D^2 based on all the characters is only slightly greater than .5766, viz., .5870. Thus little information is lost in reducing the canonical variates to two. But the distance between the two groups is almost wholly due to the difference in the values of the second, and if we wish to develop an efficient procedure for discriminating them, our best hope is to improve methods of observing variation in the second dimension.

VI. STATISTICAL CRITERION TO DETERMINE THE GROUP TO WHICH AN INDIVIDUAL BELONGS

In section V it was shown that all the groups could be adequately represented on a two-dimensional chart which was useful in establishing the relationships between the groups. This chart could also be used in considering to which of the groups an individual is most likely to belong. It would need to be divided into six regions corresponding to the six groups; and a procedure would need to be defined for assigning an individual to a particular group when he falls into the region corresponding to it. Once such a chart is made it is simple to use, and would be a valuable adjunct to any routine procedure designed for allocating men to appropriate groups (e.g., vocational selection). But the method of construction is complicated, and the reduction of a larger number of observed variables into a smaller number of canonical variables results in a loss of information unless it is strictly valid. The best method is not to use the canonical transformation, but to consider all the measurements and derive directly from them a suitable criterion for determining the group to which an individual belongs. If we have a collection of individuals belonging to various groups, a logical requirement is that the number of misclassifications should be a minimum. The general solution to this problem is given by Rao (11) and the computational aspects are given below with special reference to the present problem.

For the i th group we define the linear discriminant score

$$Li = l_1A + l_2B + l_3C - \frac{1}{2}(l_1m_1 + l_2m_2 + l_3m_3) + \log_e \pi_i,$$

where

$$l_1 = a^{11}m_1 + a^{12}m_2 + a^{13}m_3$$

$$l_2 = a^{21}m_1 + a^{22}m_2 + a^{23}m_3$$

$$l_3 = a^{31}m_1 + a^{32}m_2 + a^{33}m_3$$

m_1 , m_2 , and m_3 are the mean values of the measurements A , B , and C , π_i is the relative frequency of the members of the i th group, and a^{ij} are the elements of the reciprocal dispersion matrix in Table VIII. In the case of the group of anxiety states

$$m_1 = 2.9298, m_2 = 1.1667, m_3 = .7281 \text{ (from Table I)}$$

$$l_1 = .5432(2.9298) - .2002(1.1667) - .4208(.7281) = 1.0515$$

$$l_2 = -.2002(2.9298) + 1.7258(1.1667) + .0558(.7281) = 1.4676$$

$$l_3 = -.4208(2.9298) + .0558(1.1667) + 2.0124(.7281) = .2975$$

$$\frac{1}{2}(l_1m_1 + l_2m_2 + l_3m_3) = \frac{1}{2}\{1.0515(2.9298) + 1.4676(1.1667) + .2975(.7281)\} = 2.5047$$

$\therefore Li = 1.0515A + 1.4676B + .2975C - 2.5047 + \log_e \pi_i$; similarly the discriminant scores for the other groups are calculated and listed in Table XIV.

These discriminant scores involve the proportions of the groups in the total population, $\pi_1 \dots \pi_6$, and it is assumed that they are known accurately. If estimates derived by sampling are used in place of the true proportions, errors of estimation are introduced, the extent and effect of which have not yet been discussed from a theoretical point of view.

The present data cannot even be regarded as a representative sample of officers serving in the Army and the Navy. The sample of neurotic officers has not been exposed to any known bias, and the proportions between the numbers in the various groups may be fairly representative; but the number of normal officers is grossly under-represented. It seems impossible to obtain a reliable general estimate of the risk that a man will be referred to hospital for the treatment of a neurosis while serving in the Army or the Navy as an officer; but the indications are that even under conditions of very severe stress it is not more than 2 to 3 per cent. For proportional representation over 100 times as many normal cases should have been reported.

In the formulæ given in Table XIV we have used the proportions derived from the observed numbers in each group, although knowing that they are subject to systematic as well as chance errors. We trust that to point this out is a sufficient warning that the formulæ in this table are unsuitable for practical use or for any other purpose than to illustrate the methods. We have given them in a way which will enable more reliable estimates to be interpolated when they can be obtained.

TABLE XIV. THE LINEAR DISCRIMINANT SCORES FOR VARIOUS GROUPS

Group	Coefficients of Measurements			Constant Term	
	<i>A</i>	<i>B</i>	<i>C</i>	(a) In terms of general proportion	(b) For proportions from present data *
Normal	·2050	·1431	·1947	— ·0931 + $\log_e \pi_1$	— 1·6311
Personality change	·7204	·0649	— ·5780	— ·5107 + $\log_e \pi_2$	— 4·4465
Anxiety state	1·0515	1·4676	·2974	— 2·5047 + $\log_e \pi_3$	— 3·3137
Hysteria	1·1678	1·5679	— ·1081	— 2·7139 + $\log_e \pi_4$	— 4·7626
Psychopathy	1·3599	2·4641	·1336	— 4·9182 + $\log_e \pi_5$	— 6·9977
Obsession	1·7680	1·8611	·3573	— 5·8375 + $\log_e \pi_6$	— 8·5495

$$* \pi_1 = \cdot 21484 \quad \pi_2 = \cdot 01953 \quad \pi_3 = \cdot 44531 \quad \pi_4 = \cdot 12891 \quad \pi_5 = \cdot 12500 \quad \pi_6 = \cdot 06641$$

Given the measurements *A*, *B*, *C* of an individual we calculate the linear discriminant scores $L_1, \dots, L_6$ and assign him to the group for which his score is highest. If two scores are nearly the same he is equally likely to belong to either of the corresponding groups. If on some *a priori* information he is known not to belong to certain groups he can be assigned to one of the other groups by considering only the linear discriminant scores, as shown in Table XIV, corresponding to them.

This is the best procedure that can be laid down at present with the help of the data collected. There is no immediate prospect of obtaining accurate statistics concerning the incidence of different neurotic states in the general population or even in any particular section of it, such as the Army or the Navy. Revisions of the regulations governing recruitment and alterations in the conditions of service make it unsafe to treat the Armed Forces as forming a constant population; and similar changes are likely to affect any other section of the population defined in administrative terms. Quite apart from this, the psychiatric diagnosis is tentative and subject to revision; about some cases psychiatric opinions differ; and in the course of time methods of diagnosis change.

VII. SUMMARY

1. Data obtained by recording the incidence of a number of 'pointers' occurring significantly more frequently among neurotic officers than normal officers in the Army and the Navy have been used to illustrate methods of multivariate analysis which appear to have many potential psychological applications. These methods serve

to measure and determine the significance of overall differences between groups, taking as many dimensions into account as desired ; to determine the main orthogonal dimensions in which variation between the groups occurs ; and to calculate the likelihood that an individual may belong to one or other of the groups.

2. In their present application they show that differences exist not only between the normal and the neurotic subjects, but between subjects diagnosed as suffering from different neuroses. By far the greater part of the variation between the groups occurs in a single dimension, which might on this account be described as a general factor of neuroticism. Variations in it are not sufficient to demonstrate significant differences between the normal cases and the cases of post-traumatic personality change, or between the anxiety states and the hysterias, or between the obsessional states and the psychopathies.

3. Variations in the second dimension are below the borderline of statistical significance. They perhaps correspond to differences in the degree of instability associated with a given degree of inadequacy, and, so far as they go, appear greatest between the obsessional states and the psychopathies.

The immediate practical value of the findings is less than their value as an illustration of the possibilities afforded by the techniques used. We wish to express our thanks to Professor M. S. Bartlett for his helpful criticisms.

REFERENCES

1. Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* XXXIV, 33.
2. Bartlett, M. S. (1948). Internal and external factor analysis. *Brit. J. Psych. (Statistical Section)*, 1, 73.
3. Burt, Sir Cyril (1938). The analysis of temperament. *Brit. J. Med. Psych.*, XVII, 158.
4. Eysenck, H. J. (1944). Types of personality—a factorial study of 700 neurotics. *J. Ment. Sci.*
5. Eysenck, H. J., et al. (1947). *Dimensions of Personality*. Kegan Paul, Trench, Trubner and Co.
6. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen. Lond.*, VII, 179.
7. Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugen. Lond.*, VIII, 376.
8. Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear regression. *Ann. Eugen. Lond.*, IX, 238.
9. Hotelling, H. (1931). The generalization of student's ratio. *Ann. Math. Stat.*, II, 360.
10. Mayer Gross, W., Moore, J. N. P., and Slater, P. Forecasting the incidence of neurosis in officers of the Army and the Navy. *J. Ment. Sci.* (in press).
11. Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. Paper read before the Royal Statistical Society, April 6th.
12. Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika*, XXXV, 58.
13. Slater, Eliot (1943). The neurotic constitution. *J. Neurol. and Psychiat.*, VI, 1.
14. Slater, E., and Slater, P. (1944). A heuristic theory of neurosis. *J. Neurol. and Psychiat.*, VII, 1 and 2.
15. Slater, Patrick (1947). The factorial analysis of a matrix of 2×2 tables. *J. Roy. Stat. Soc.*, X, 1 and 2 (Supplement).
16. Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, XXIV, 471.

THE CONCEPT OF EQUIVALENT SCORES IN SIMILAR TESTS

By PHILIP D. GREENALL

Air Ministry (Science 4)

I. Introduction. II. The Regression Lines. III. Likely Single Lines. IV. Consequences of Re-scaling Scores. V. Symmetry in Errors of Estimation. VI. Considerations of Cumulative Frequency. VII. Conclusions and Summary.

I. INTRODUCTION

A psychological or educational test may have been in use for some time to measure an aptitude or attainment, and its replacement by another test, intended to measure the same sort of aptitude or attainment, may have become desirable. Alternatively one test may have been in use in one environment, and a different test, designed to measure the same feature, in another. In both of these cases it sometimes happens that norms, pass-marks for various purposes, weighting formulæ or profiles (if the test is one of a selection battery), and directions for use are available, expressed in terms of scores in one of the tests but not in the other. Approximate figures relating to the second test may be desired. Or there may be some reason for wishing to compare some of the persons assessed only by one test with other persons tested only by the other. In circumstances such as these it might be useful if a mutual conversion table (or graph, or equation) could be devised to relate scores in the two tests. This would lay down the pairs of scores in the two tests, which, on fairly reasonable grounds, could be regarded as roughly and mutually equivalent to each other. Mutuality is emphasized because a single table that could be read both ways would be desirable in practice.

Those aspects of the problem that are dealt with here were considered by the author at first in ignorance of previous work bearing on the subject. In this fresh investigation a different approach was adopted from those found in the scattered earlier publications. K. Pearson (1901) discussed a purely mathematical problem bearing on this one. A problem in medical statistics resembling our own interested M. Greenwood and G. U. Yule (1915). In the statistics of economics a similar one has interested a few recent writers, starting with R. Frisch (1934). More recent related discussions, in the purely mathematical sphere, are due to C. F. Roos (1937), A. Wald (1940), and R. C. Geary (1942). In the field of psychology and education, the matter was discussed shortly before the first world war by a British Association Committee on Factors in Education (1911) which included J. A. Green and C. Burt as secretaries. Apparently the earliest paper actually published on the subject is one by A. Otis (1922), whose conclusions agree with those of the present writer. Suggestions similar to Otis's, based on the mutual equivalence of scores in standard measure, have been put forward by various writers, e.g., Burt (1917) and Thomson (1928, 1945). This paper therefore offers a discussion of certain theoretical aspects of the problem which do not seem to have been dealt with explicitly in the previous publications on account of their different approaches.

From the theoretical point of view it will frequently be convenient to forget that scores in two similar tests are being considered, and to discuss just a pair of correlated variable numbers x and y . If we care to treat x and y as being continuous variables (that is, variables able to assume all possible decimal values between the integral values), we can reconcile this treatment with the idea of test scores by assuming that

the ability measured by the tests can vary continuously, but the measures are scored or grouped to the nearest integer for convenience. Similarly, if we find it convenient for any purpose to regard x and y as capable of assuming positive or negative values as large as desired without limit, while real test scores lie within an arbitrarily determined range, the discrepancy will not be serious provided the possible range of test scores is already so extensive that scores even approaching its limits (say, zero and 100) are fairly improbable, as when the graphical representation of the distribution of scores is roughly bell-shaped.

If, in the theoretical discussion, we speak on occasion as if the pairs of numbers (x, y) are able to assume an infinite number of different possible pairs of values, each pair with its own relative probability of occurring, this need not clash with the common sense knowledge that only a finite number of persons has ever sat for any pair of real tests. Any finite set of pairs of scores (x, y) may be regarded as forming just one sample from the conceptual infinite set of all the pairs that could possibly have occurred in like circumstances. Certain numbers that can be calculated for any sample (for example, the mean of the x 's in the sample, or the correlation coefficient between x and y in the sample) are conveniently regarded as estimates of the numbers that in imagination it would be possible to calculate similarly for the whole of the conceptual infinite population of pairs from which that sample can be regarded as having been drawn. The distribution of x and y in this conceptual infinite population will determine the relative probability of any particular pair of values (x, y) turning up in a real sample. Numbers, like the mean and the correlation coefficient, that describe the distribution of the conceptual infinite number-population, will therefore determine the relative probabilities of the pairs of numbers found in real samples. These fundamental numbers are the *parameters* of the distribution of the conceptual infinite population (or of the distribution of relative probabilities of the possible pairs (x, y)), while the corresponding figures (mean, correlation, etc.), calculated for each real, finite sample, are among the *statistics* of that sample. *Statistics* that provide estimates of the conceptual true *parameters* enable us to make predictions about further samples not yet measured or tested. Before mathematical techniques can be used, we must always construct such convenient and often slightly idealized mathematical models of the actual state of affairs. In the models, mathematical quantities which we can manipulate must bear to each other relations similar to the relations observed between the corresponding real things. To describe our mathematical model in this case has been the purpose of the discussion in these last two paragraphs, which show that the customary statistical notions mentioned are applicable to test scores.

So far as notation is concerned, it is customary and convenient to use small Roman letters to denote statistics calculated from a real finite sample (r for the correlation, say m_x for the average x , s_y for the standard deviation of the y 's, and so on), and the corresponding small Greek letters for the corresponding parameters of the conceptual distribution (ρ for the correlation, μ_x for the mean x , σ_y for the standard deviation of the y 's, and so on). Therefore ρ is not to be read as Spearman's rank order correlation. A convenient symbol, which will be used in what follows, is that for the so-called 'expected' value of a variable quantity; roughly, this is the mean of all the values that would occur in infinitely large samples. (It can be calculated by summing the product of each value by the chance of its occurring expressed as a fraction.) That the 'expected' value of x is the mean of x can be written in terms of this convenient symbolism, $E(x) = \mu_x$. Similarly, since σ_y^2 roughly means the average of all the squared deviations from the mean y (i.e., from μ_y), we have, as a definition of standard deviation,

$$\sigma_y = \sqrt{E[(y - E(y))^2]}. \quad \text{And similarly } \rho = E[(x - \mu_x)(y - \mu_y)] / \sigma_x \sigma_y.$$

The correlation coefficient between two variables can be regarded as a measure of the approach to perfect linearity in the relation between them: loosely speaking, of how

closely the points (x, y) on a graph cluster about a suitable straight line. Now the original problem, concerning test scores, involved similar tests both designed to measure the same sort of ability. We shall confine our attention to the case when sets of scores on the particular tests concerned correlate quite highly. The problem is not likely even to be of interest in the case of tests that do not correlate quite highly, and so only the former state of affairs will be considered here. Hence, in our mathematical model, points that represent graphically the pairs of variables (x, y) will mostly lie close to some straight line. The 'expected' value of the distance (or of the square of the distance) of (x, y) from such a line will be small. The distance that can of course be measured parallel to the y -axis, or to the x -axis, or perpendicularly to the line itself, or in any other stipulated direction. We shall restrict the scope of this paper to deciding which, of several possible straight lines that fit the 'scatter' of points closely, it would be most reasonable to use to represent the general relation between x and y . It will be possible to use the line chosen, or its equation, to construct our required conversion table giving values of x and y that can be treated as mutually equivalent for most practical purposes.

II. THE REGRESSION LINES

As a first guess we might suggest using a regression line.

That for y on x has the equation : $y - \mu_y = \frac{\sigma_y}{\sigma_x} \rho (x - \mu_x)$,

and that for x on y has the equation : $y - \mu_y = \frac{\sigma_y}{\sigma_x} \frac{1}{\rho} (x - \mu_x)$.

Let us consider the first of these. It is a straight line whose position is uniquely defined for each distribution, however the points (x, y) are distributed and whatever the correlation ρ . If the distance of a point (x, y) from a line L (distance being measured here parallel to the y -axis) be denoted by 'dist. $((x, y) \text{ to } L)$,' then for all points (x, y) the 'expected' or average squared distance will be $E[(\text{dist. } ((x, y) \text{ to } L))^2]$. The regression line for y on x is merely that line for which this 'expected' value is less than for any other straight line. Hence we can regard the line in another light. If we know only one variable of a pair (x, y) (e.g., the score of a certain candidate on one test, but not on the other), we may wish to estimate y . Corresponding to x will be a value of y , Y , say, on this regression line. If we take Y as an estimate of y , y being the true value corresponding to x in (x, y) , our estimate will be in error by an amount $(y - Y)$. The square of our error will be $(y - Y)^2$. Estimating for all possible (x, y) 's, our 'expected' or average squared error of estimation will be $E[(y - Y)^2]$. The regression line is the straight line that makes this expectation a minimum. Similar remarks apply, *mutatis mutandis*, to the regression line for x on y , in this case distances being measured parallel to the x -axis.

But, for our purpose, the regression lines (or equations) are unsuitable simply because they are a *pair* of lines. We stipulated that we sought a mutual conversion table, giving mutually equivalent values of x and y . If we used the regression lines, that for y on x would give us Y , say, as the value of y equivalent to X . But, to find the value of x equivalent to this value Y , we should have to use the other line (for x on y), and this would give a value X' , say, different from the original X . So the equivalence established would not be mutual.

III. LIKELY SINGLE LINES

Before proceeding to consider the relative merits of various single straight lines, we may slightly simplify our search and our algebra. Consider any two parallel straight lines on the graph, one L through the mean point or centroid (μ_x, μ_y) of the distribution and the other M parallel to L , but not passing through the centroid. Then, if we consider distances from points (x, y) to L and to M , all measured in *any* stipulated direction, the 'expected' value of these distances squared will be less to L than to any line M parallel to L . Therefore we shall direct our attention to straight lines passing through the centroid (μ_x, μ_y) of the bivariate distribution, since they 'fit best.' Hence, it will be simpler to consider all the x 's reduced by an amount μ_x , and all the y 's by μ_y . This will make our new means both zero, and the centroid of the distribution will be $(0, 0)$, the origin of our graph.

Certain alternative single straight lines at once suggest themselves as possibilities for our purpose, and a comparison of their relative merits has to be made. One is the 'line of closest fit.' This line has been discussed before (Karl Pearson, 1901), but not from this particular viewpoint. It is the line which fits the distribution 'closest' in the sense that the 'expected' or average value of the squared distances from the points (x, y) to the line is less than for any other straight line, distances from the line being measured in this case perpendicularly to the line itself (and not parallel to one or other of the axes of the graph, as for the regression lines). It can be shown that the major axis of the so-called correlation ellipse, in cases for which it is defined, coincides with this line; and that, whatever the distribution of pairs of values (x, y) , the line passes through the centroid and makes with the x -axis an angle A such that

$$\tan 2A = \frac{2\rho \sigma_x \sigma_y}{\sigma_x^2 - \sigma_y^2}.$$

Another straight line which also has an immediate intuitive appeal is the one which would arise were equal 'standardized' variables ('standardized' scores on the two tests) deemed to be mutually equivalent. This is the line recommended on certain grounds by Otis in 1922. He called it the 'relation line.' Before reading Otis's paper I had called it the 'equivalence line.' The latter term is perhaps more descriptive for the particular circumstances under discussion here, while the former would be of more general application and could be used also in cases where x and y were not merely different measures of roughly the same characteristic. The equation of this line is :

$$y = \frac{\sigma_y}{\sigma_x} x.$$

Whatever the distribution, this line passes through the centroid, and makes with the x -axis an angle B such that

$$\tan B = \frac{\sigma_y}{\sigma_x}; \text{ i.e., } \tan 2B = \frac{2\sigma_x \sigma_y}{\sigma_x^2 - \sigma_y^2}.$$

These two lines, of 'closest fit' and of 'equivalence,' can coincide only if $\tan 2A = \tan 2B$. This can be the case only if $\rho=1$ (giving perfect correlation), or if $\sigma_x = \sigma_y$, in which latter case angles A and B are both 45° .

A choice between various apparently eligible lines should become possible if we can state some reasonable requirements which point to one rather than another.

The choice of requirements must necessarily be to some extent arbitrary. Those suggested in this paper are three which are likely to seem reasonable in psychological and educational work, especially in connexion with the particular problem originally outlined.

As mentioned earlier, the discussion that follows does not adopt the approach of writers already mentioned, such as Otis, Wald, and Geary. Their viewpoint would entail our regarding the two tests as intrinsically inaccurate measures of two ideal quantities, between which there subsisted a perfect linear relation obscured only by the errors. This would include, as a special case, that of two *inaccurate* measures of just one underlying ideal quantity. Quite a number of mathematical assumptions would have to be made concerning these errors of measurement. Otis came to the same conclusion as we shall presently reach by our own route, by seeking the equation of a straight line relating the ideal underlying error-free measurements. Both Wald and Geary were concerned with the following mathematical problem. If the conceptual straight line relating the conceptual error-free variables X and Y has the equation

$$Y = mX + c,$$

then what are the values of m and c ? If we followed Geary we should consider the problem of calculating m and c in terms of the parameters of the bivariate frequency distribution of the two observed variables in which the errors are intrinsic. If his original assumptions are valid, various values of m and c can be found expressed in terms of the so-called cumulants (or semi-invariants) of this distribution. Using Wald's approach, however, we should consider the problem of estimating m and c in terms of the statistics derived from a finite sample of N pairs of the observed variables which involve error. If his assumptions hold good, it can be shown that the statistics he gives provide unbiased estimates of m and c (which means that the 'expected' values of his formulæ are m and c respectively). Interesting though a knowledge or estimate can be of the true relation obtaining between the conceptual error-free variables, the practical usefulness of this knowledge is liable to be diminished by the fact that these ideal underlying quantities themselves remain unknown. In actuality we can deal only with the observed values, from which the alleged errors of measurement are inseparable. If we did wish to regard from this viewpoint our problem of two tests of the same ability, we should have to postulate an underlying absolute and natural measure, Z say, of this ability. The observed measures x and y could both be regarded as being compounded of Z and random errors (or, if one of the marking scales involved a systematic distortion of the natural measure, as being compounded of a function of Z with random errors). However, our treatment will seek a straight line relating the *observed* and not the conceptual error-free scores, and the requirements stipulated will have a utilitarian bias for this reason.

IV. CONSEQUENCES OF RE-SCALING SCORES

It would not be unreasonable to require the following property in a line used as a basis for the construction of mutual conversion tables. If two variables (test scores) were all multiplied, the first by one constant positive factor, the second by another (all the x 's by a , say, and all the y 's by b), the new line chosen by the same rule should determine a revised conversion table, which would pair off aX with bY , whenever X and Y were paired off in the original table before multiplication. In other words, the equivalence set up should be independent of mere linear changes of scale.

Let us see if the 'equivalence line' suggested does in fact satisfy this criterion. If X was originally deemed the mutual equivalent of Y , because the point (X, Y) lay on the equivalence line, then :

$$Y = \frac{\sigma_y}{\sigma_x} X.$$

But, after re-scaling (multiplying by positive numbers a and b), the new standard deviations of the two new variables will be $a\sigma_x$ and $b\sigma_y$. So the same principle will lead to a new equivalence line

$$y = \frac{b\sigma_y}{a\sigma_x} x$$

to be used for constructing the new conversion table. If now we take the corresponding value aX of the re-scaled variable ax , the ' by ' value, Y' say, given by the new line will be :

$$Y' = \frac{b\sigma_y}{a\sigma_x}(aX) = b\left(\frac{\sigma_y}{\sigma_x}X\right) = bY,$$

which is the value hoped for.

So the equivalence line does satisfy a very reasonable requirement. However, this latter is not a distinguishing property, for any other line with an equation of the general form,

$$y = k \frac{\sigma_y}{\sigma_x} x$$

(for any constant k), will do the same thing.

It is easy to show that Pearson's ' line of closest fit ' does not have this property. We showed at the end of the paragraph before last that the ' line of closest fit ' and the ' equivalence line ' coincide if the standard deviations σ_x and σ_y are equal. Suppose they are equal to start with, ensuring coincidence, and that ρ is not equal to unity. Now re-scale the variables, multiplying all the x 's by a , and all the y 's by b (a and b both being positive but not equal). The formulæ given for $\tan 2A$ and $\tan 2B$ will no longer be equal. Hence the angles A and B , which the lines make respectively with the x -axis, will no longer in general be equal. Hence the two lines will now diverge. But we showed in the last paragraph that the equivalence line will have assumed a new position, which is the same, relative to the new points (ax, by) , as was its former position relative to the original points. Therefore a line which coincided with it initially, but now diverges from it, cannot behave likewise after the re-scaling. Hence, from this point of view, Pearson's ' line of closest fit ' would be unsuitable as the basis of a mutual conversion table. For example, it would obviously be undesirable if our first conversion table showed $y=30$, say, as the equivalent of $x=50$, while, after the x 's had all been perhaps doubled, to give a new variable $x'=2x$, our new line gave a table which did not give $y=30$ as the equivalent also of $x'=100$.

V. SYMMETRY IN ERRORS OF ESTIMATION

So far the arguments advanced have had rather a negative bias. We have been finding shortcomings in lines that were *prima facie* possibilities for our purpose. We have not yet suggested requirements for such lines that will perhaps single out the ' equivalence line ' as being uniquely suitable, in addition to having its intuitive appeal through ascribing equivalence to equal ' standardized ' variables. Therefore let us consider another reasonable requirement.

We have already discussed the definition or characteristic property of the regression line for y on x . It is the straight line for which the ' expected,' or average, value of the squared error is a minimum if the line is used for estimating y , in the pair (x, y) , when we are given only the value of x . (Error in this case is measured parallel to the y -axis.) It can be shown fairly easily that the ' expected ' value of this squared error of estimation of y is $\sigma_y^2(1 - \rho^2)$. Similarly, if we use the other regression line, that for x on y , to estimate the x 's of pairs (x, y) , when we are given only the y 's, errors must be measured now parallel to the x -axis. In this case, the ' expected ' value of the squared error of estimation of x is $\sigma_x^2(1 - \rho^2)$. Both of these error expectations, the former belonging to the line used when we are estimating y 's, knowing only the x 's, and the second to the line for estimating x 's, knowing only the y 's, are minima. Any single straight line, therefore, with which we replace the pair of regression lines, will give inferior results (except in the trivial case when the lines coincide because $\rho=1$). This is obviously so, as at least one of the expectations of error must be increased above its corresponding minimum value for which we have given the formula. To preserve symmetry between our treatments of x and y , in the absence of any good reason for treating one variable

more favourably than the other, it would be reasonable to require that, when a single estimation line is chosen, the following should be the case. The proportional increase in the 'expected' squared error of estimation of y , as compared with its ideal minimum value of $\sigma_y^2(1-\rho^2)$, should be the same as the proportional increase in the 'expected' squared error of estimation of x , as compared with its ideal minimum value of $\sigma_x^2(1-\rho^2)$.

To determine which straight line through the centroid has this desirable property, let us assume the line sought has an equation involving an unknown slope k : $y = kx$. We shall attempt to evaluate the unknown constant k .

Consider one point (X, Y) representing a pair of variables (scores of the same candidate on two tests). If we know X only, and use the line $y = kx$ to estimate Y , the error will be $(Y - kX)$, which will not be zero unless the true (X, Y) really lies on the line. So the 'expected' squared error of estimation of y , which could be abbreviated to 'Error²(y, k)', is given by:

$$\begin{aligned}\text{Error}^2(y, k) &= E[(Y - kX)^2] \\ &= E(Y^2 - 2kXY + k^2X^2) \\ &= E(Y^2) - 2kE(XY) + k^2E(X^2).\end{aligned}$$

But, since we are now measuring all our variables from a centroid placed for convenience at the origin $(0, 0)$ (i.e., $\mu_x = E(X) = \mu_y = E(Y) = 0$), certain definitions simplify to give:

$$\sigma_y^2 = E(Y^2), \text{ and } \sigma_x^2 = E(X^2),$$

and

$$\frac{E(XY)}{\sigma_x \sigma_y} = \rho.$$

Hence

$$\text{Error}^2(y, k) = \sigma_y^2 - 2k \sigma_x \sigma_y \rho + k^2 \sigma_x^2.$$

Suppose now we know only Y of the pair (X, Y) , and wish to estimate X using the same line. Let us rewrite the equation of the line as

$$x = \frac{1}{k}y.$$

The 'expected' squared error of estimation of x , which could be abbreviated to

$$\begin{aligned}\text{'Error}^2(x, k)\text{' } &= E[(X - \frac{1}{k}Y)^2] \\ &= \sigma_x^2 - \frac{2}{k} \cdot \sigma_x \cdot \sigma_y \cdot \rho + \frac{1}{k^2} \cdot \sigma_y^2 \text{ (for reasons similar to those holding in the case of } y) \\ &= \frac{1}{k^2}(\sigma_y^2 - 2k \cdot \sigma_x \cdot \sigma_y \cdot \rho + k^2 \sigma_x^2) \\ &= \frac{1}{k^2} \text{Error}^2(y, k).\end{aligned}$$

We required that the use of the single line should involve the same increases, proportionately, on the two ideal minimum values that arise when the separate regression lines are used:

$$\begin{aligned}\text{i.e., } \frac{\text{Error}^2(x, k)}{\sigma_x^2(1-\rho^2)} &= \frac{\text{Error}^2(y, k)}{\sigma_y^2(1-\rho^2)}; \\ \text{hence } \frac{\sigma_y^2(1-\rho^2)}{\sigma_x^2(1-\rho^2)} &= \frac{\text{Error}^2(y, k)}{\text{Error}^2(x, k)}, \\ \text{and hence } \frac{\sigma_y^2}{\sigma_x^2} &= k^2 \text{ (provided } \rho \neq \pm 1).\end{aligned}$$

If ρ is positive, these expectations will be less if k is positive than if k is negative. Hence, taking the positive roots, we get $k = \frac{\sigma_y}{\sigma_x}$, giving us the equivalence line, $y = \frac{\sigma_y}{\sigma_x}x$. Such a requirement of symmetry thus selects this line positively and uniquely.

It would be interesting to see just how large is this common deterioration in the accuracy of estimation when a single line is used instead of the ideal pair.

$$\begin{aligned}\text{Error}^2(y, k) &= \sigma_y^2 - 2k \cdot \sigma_x \cdot \sigma_y \cdot \rho + k^2 \sigma_x^2 \\ &= 2\sigma_y^2 - 2\sigma_y^2 \rho \quad \left(\text{since } k = \frac{\sigma_y}{\sigma_x} \right) \\ &= \sigma_y^2 \cdot 2(1 - \rho).\end{aligned}$$

Using the appropriate regression line we should have instead: $\sigma_y^2(1 - \rho^2)$, i.e., $\sigma_y^2(1 + \rho)(1 - \rho)$. So we have increased our minimum 'expected' squared error of estimation of y by a factor which is the ratio of these quantities, i.e., by $\frac{2}{1 + \rho}$; and similarly for estimating x . If we turn the 'expectations' into quantities comparable with standard deviations, by considering their square roots, this factor becomes $\sqrt{\frac{2}{1 + \rho}}$. When ρ exceeds 0.7, the value of $\sqrt{\frac{2}{1 + \rho}}$ is less than 1.085, and the 'standard deviations' of errors of estimate are increased by less than 8½% through our using the single equivalence line instead of the ideal pair.

VI. CONSIDERATIONS OF CUMULATIVE FREQUENCY

Hitherto, the results obtained have been general and quite independent of the form of the (x, y) distribution. Even an assumption of continuity in the variables would have been redundant before this stage, since the arguments advanced fit the case of discrete integral values also. The assumption of unlimited x and y ranges, to 'plus and minus infinity,' would also have been redundant, as possible limits to x and y have not hampered our arguments. Further, arguments of a similar form could have been applied as easily with changes of notation to finite samples as to the hypothetical infinity of possible pairs (x, y) . However, in considering the next point (especially when we come to the normal distribution, to which approximate so many distributions found in practice), the mathematical model involving continuous variables of unlimited range will be useful.

For continuous variables, a very reasonable definition of mutual equivalence between pairs of scores on similar tests of the same sort of ability would be one that paired off X with Y if the proportion of x -scores less than X were equal to the proportion of y -scores less than Y . On this definition, equivalent scores would, for the same population, give rise to the same percentiles, the same deciles, the same quartiles, or in general to the same 'quantiles,' to use the general term for all 'partition values' or cumulative measures of relative ranking. Such a definition would be especially useful in cases in which there was obviously a close functional relation between the two measures x and y , which, however, was more difficult to treat because the points (x, y) were more closely fitted by a curve than a straight line. This usefulness does not preclude the approach being a sound one worth investigating also for the circumstances under consideration here, in which, graphically speaking, a straight line fits the data tolerably well.

At any point (x, y) , the relative frequency with which points in that neighbourhood will turn up in random sampling can be denoted by the symbol $F(x, y)$, say. This is the 'probability density function' for the pairs of values (x, y) . Since it is certain (i.e., the probability equals unity) that any random pair of scores (x, y) will lie *somewhere*, we must have in the notation of the integral calculus:

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x, y) \cdot dy \cdot dx.$$

The Concept of Equivalent Scores

The proportion of x -scores less than X (i.e., the probability that, if we select at random point (x, y) , we shall find $x < X$) is given by the expression :

$$\int_{-\infty}^X \int_{-\infty}^{\infty} F(x, y) \cdot dy \cdot dx .$$

Similarly the proportion of y -scores less than Y is given by $\int_{-\infty}^Y \int_{-\infty}^{\infty} F(x, y) \cdot dx \cdot dy$.

These proportions will be equal if and only if the two latter double integrals are equal. The relation between X and Y , obtained by equating these expressions, will not necessarily always represent a straight line as preferred, but possibly a curve, since we have so far placed no restriction on $F(x, y)$, which might describe some extremely inconvenient distribution. The possibility of a straight line resulting must now be investigated.

Let us consider the possibility of a relative frequency (or probability density) function that is 'symmetrical' in the two variables, in the sense that its value remains unchanged if they change places with one another, provided only a suitable rescaling takes place to allow for arbitrary differences in their scales. More explicitly, suppose there exists a probability density function $F(x, y)$ having the property that there is some positive constant c for which

$$F(x, y) \equiv F\left(\frac{y}{c}, cx\right) .$$

To simplify our algebraic expressions, we are still assuming x and y are measured from their means, giving $E(x) = E(y) = 0$, and the previous simplified definition $\sigma_x^2 = E(x^2)$.

Hence

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 F(x, y) \cdot dy \cdot dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 F\left(\frac{y}{c}, cx\right) \cdot dy \cdot dx . \end{aligned}$$

This last integral can be transformed by the simple substitutions $x = z/c$ and $y = cw$, which necessitate the replacement of ' dx ' by ' dz/c ' and ' dy ' by ' $c \cdot dw$.' Hence :

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{z}{c}\right)^2 F(w, z) c \frac{1}{c} dw \cdot dz \\ &= \frac{1}{c^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^2 F(w, z) dw \cdot dz = \frac{1}{c^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 F(x, y) \cdot dx \cdot dy \\ &= \frac{1}{c^2} E(y^2) = \frac{1}{c^2} \sigma_y^2 . \end{aligned}$$

Taking positive roots throughout gives : $c = \frac{\sigma_y}{\sigma_x}$.

Let us now revert to the two integrals whose equality we sought. If the proportion of x -scores less than X equals the proportion of y -scores less than Y , we have :

$$\int_{-\infty}^X \int_{-\infty}^{\infty} F(x, y) dy \cdot dx = \int_{-\infty}^Y \int_{-\infty}^{\infty} F(x, y) \cdot dx \cdot dy .$$

But, if we make the same substitutions as before in the left-hand side of this equality, and note that when $x = X$, $z = cX$, we get :

$$\begin{aligned} \int_{-\infty}^X \int_{-\infty}^{\infty} F(xy) dy \cdot dx &= \int_{-\infty}^{cX} \int_{-\infty}^{\infty} F\left(\frac{z}{c}, cw\right) c \frac{1}{c} \cdot dw \cdot dz \\ &= \int_{-\infty}^{cX} \int_{-\infty}^{\infty} F(w, z) \cdot dw \cdot dz = \int_{-\infty}^{cX} \int_{-\infty}^{\infty} F(x, y) \cdot dx \cdot dy. \end{aligned}$$

This latter expression will be equal to the right-hand side of our original equation in integrals if and only if the upper limits of the two outer integral signs are equal ; that is, if and only if $Y = cX$, i.e., $Y = \frac{\sigma_y}{\sigma_x} X$;

that is, if and only if (X, Y) lies on the equivalence line $y = \frac{\sigma_y}{\sigma_x} x$.

It remains only for us to decide whether bivariate distribution functions, possessing the sort of symmetry which we stipulated in the last paragraph for $F(x, y)$, are likely to be found in practice. The answer is that the bivariate normal distribution, to which so many real distributions do approximate quite closely, is one example of such a function. In this case $F(x, y)$ means

$$\frac{1}{2\pi \cdot \sigma_x \cdot \sigma_y \cdot \sqrt{1-\rho^2}} \cdot \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2} \right) \right].$$

In the above expression, if we substitute y/c for x and cx for y (where $c = \frac{\sigma_y}{\sigma_x}$, as discovered in the last paragraph) we get exactly the original expression. So, in the case of a certain class of frequency distributions, including the normal one, the equivalence line does pair off scores X and Y in our conversion table in such a way that the proportion of candidates scoring less than X on one test will equal the proportion scoring less than Y on the other. In general our stipulations cannot hold even approximately unless x and y separately are distributed fairly similarly : for instance, unimodal distributions of x and y with markedly opposite skews are unsuitable.

VII. CONCLUSIONS AND SUMMARY

For the theoretical reasons set out in this paper, as well as for those advanced previously by Otis, it appears to be a reasonable and useful suggestion that equal standardized scores on similar tests should be deemed mutually equivalent.

The arguments put forward here can be summarized as follows :

(1) If there is a fairly high correlation coefficient between scores on two tests intended to measure the same ability, it seems reasonable to seek a single straight line (or linear equation) to determine before drawing up a conversion table the pairs of scores to be regarded as roughly mutually equivalent.

(2) If we wish to estimate an individual's score on the second test, knowing only his score on the first, or vice versa, the correct regression line provides better estimates than any other straight line. The regression lines are rejected for our present purpose because they involve a *pair* of lines, and so the equivalences they yield cannot be mutual.

(3) Previous papers have discussed the case of a single underlying straight line relating conceptual error-free measures. This paper deals only with observed scores.

(4) *Prima facie* it would appear that at least two different single straight lines are suitable for use in drawing up a conversion table. Pearson's 'line of closest fit' by definition fits the data more closely than does any other straight line, a fact which recommends its use. At the same time the line which pairs off scores which become equal to each other when 'standardized,' referred to here as the 'equivalence line,' also seems a likely possibility. Arbitrary but useful requirements must be laid down before a line can be chosen.

(5) If we require our method of pairing off scores on the two tests to be proof against mere linear changes of scale, we can accept the equivalence line, but must reject Pearson's 'line of closest fit.' This is because it can be shown that, after such a re-scaling, the new version of the equivalence line bears the same relation to the new pairs of scores, as did the original version to the original pairs, while, graphically speaking, in general the 'line of closest fit' assumes a different relative position.

(6) In replacing the regression lines by a single line, we could try to maintain symmetry of treatment between the two variables. Hence we seek a single line, estimation by means of which will magnify the minimum 'expectations' of error in the same proportion for each variable. The equivalence line is the only line which satisfies this requirement.

(7) When we seek a relation that will deem a pair of scores X and Y mutually equivalent if and only if the proportion of x -scores less than X equals the proportion of y -scores less than Y , we aim at pairing off scores that give rise to equal percentile ranks, or, in general, equal quantiles. In the case of continuous bivariate distributions which satisfy a certain simple condition (which is satisfied by the normal distribution amongst others), only the equivalence line will provide this relation.

(8) Therefore, the appeal of the equivalence line appears to be justified. For non-zero means, its full equation is $\frac{y - \mu_y}{\sigma_y} = \frac{x - \mu_x}{\sigma_x}$, but, for standardized scores, simply $y = x$. In the conditions described, there seems no theoretical objection to using it for drawing up conversion tables to give pairs of roughly and mutually equivalent scores.

REFERENCES

1. Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 6th Series, II, pp. 559-72.
2. Greenwood, M., and Yule, G. U. (1915). The statistics of anti-typhoid and anti-cholera inoculations and the interpretation of such statistics in general. *Proc. Roy. Soc. Med.* (Section of Epidemiology and State Medicine), VIII, pp. 113-94 (especially pp. 150-1).
3. Burt, C. (1917). *Distribution of Educational Abilities*, p. 14; cf. *Mental and Scholastic Tests* (2nd ed.), pp. 439f. and refs.
4. Otis, A. (1922). The method of finding the correspondence between the scores in two tests. *J. Educ. Psychol.*, XIII, pp. 529-44.
5. Thomson, G. H. (1928). The mental age concept and standardization of group tests. *Psychol. Rev.*, XXXV, pp. 398-413 (especially pp. 400-1).
6. Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems*. Universitets Økonomiske Institutt, Oslo.
7. Roos, C. F. (1937). A general invariant criterion of fit for lines and planes when all variates are subject to error. *Metron*, XIII (1), pp. 3-20.
8. Wald, A. (1940). The fitting of straight lines when both variables are subject to error. *Ann. Math. Stats.*, XI, pp. 284-300.
9. Geary, R. C. (1942). Inherent relations between random variables. *Proc. Roy. Irish Acad.*, XLVII (A), pp. 63-76.
10. Thomson, G. H. (1945). The distribution of intelligence among university and college students. *Brit. J. Educ. Psychol.*, XV, pp. 76-9 (especially p. 78).

SUBDIVIDED FACTORS

By CYRIL BURT

Department of Psychology, University College, London

I. *The Hierarchical Arrangement of Factors.* II. *Demonstration from Body-Measurements : Data.* III. *Bipolar Analysis : (a) Analysis by General Factors ; (b) Analysis by Subdivided Factors.* IV. *Group Factor Analysis with Subdivided Factors.* V. *Correlations with Temperamental Factors.* VI. *Summary and Conclusions.* VII. *Appendix : A Shortened Method of Factor Analysis.*

I. THE HIERARCHICAL ARRANGEMENT OF FACTORS

The Structure of the Mind. In its earliest phases the chief purpose of factor analysis was to facilitate the classification of persons according to what Karl Pearson termed 'index characters,' that is, in psychology, according to a diagnostic scheme of mental traits or types. At the same time it was hoped that the results might also throw some light on the structure of human personality, particularly in its psychological aspects. During recent years, however, the latter problem has dropped into the background. Those who publish lists of mental factors are content merely to establish the existence of certain 'primary abilities' or 'unitary traits,' without troubling to consider their place in the organization of the mind or to fit them into any kind of system. Indeed, some writers have suggested that, to a first approximation at least, the mind can be assumed to function by processes equivalent to random sampling.

This is very different from the conceptions that inspired the first attempts at factorizing mental performances. However divergent their actual conclusions, they were nearly always based on some rational hypothesis that had its roots in the current theories of the day. During the first two decades of the present century, when factorial methods had their origin, there was a growing conviction among British psychologists that the mind, like the nervous system, was organized into a 'hierarchy of levels.'¹ This fundamental notion was accepted by nearly all the rival schools. But in their endeavours to verify its consequences by statistical means the different investigators were sharply divided about the way it should be interpreted.

Alternative Views. In accordance with the general trends prevailing in philosophy and science about that time, the more theoretical writers usually inclined towards a *monistic* interpretation, the more empirical towards a *pluralistic*. Thus, those who were interested primarily in *general* psychology tended to assume that mental processes on the higher levels differed from those on the lower levels merely in complexity or degree, not in quality or kind. This doctrine was a reaction against the exaggerated emphasis on discontinuous qualitative differences, which formed a leading feature of the older faculty school. It was introduced by the associationists and retained by their opponents. Spencer, for example, held that all conscious processes had progressively evolved from the same relatively simple

¹ In psychology the introduction of this phrase is due to McDougall (cf., for example, his *Physiological Psychology*, 1905, p. 23, and his papers on 'Physiological Factors in the Attention Process,' *Mind*, N.S., XI, 1902, pp. 333). The idea of a 'neural hierarchy' with three or more main 'levels' is due to Hughlings Jackson ('On the Relations of the Divisions of the Nervous System to One Another,' *Brit. Med. Journ.*, 1898, pp. 65f. ; cf. Sherrington, *Integrative Action of the Nervous System*, 1906, p. 314). See also Allport, G. W., *Personality*, pp. 45, 139, and refs.

cognitive activity, and that the successive levels were simply an expression of the 'increasing differentiation and integration' of one essential mental function which he termed 'intelligence.' Sully and Ward were equally agreed on the existence of a single universal cognitive process, which Sully called 'discrimination' and Ward 'attention.'¹

On the other hand, those who were interested rather in *individual* psychology, like Galton (1) and Binet (2), still accepted the traditional assumption of qualitatively distinct mental faculties: some of these they considered to be relatively wide and comprehensive, like the capacity which Galton called 'natural ability' and Binet termed 'natural' or 'general intelligence';² others, like discrimination, memory, imagination, etc., which Binet termed 'partial aptitudes,' were considered to have a much narrower range. Investigations demonstrating individual differences in imagery, and studies of special disabilities among school children, seemed strongly to support this view. Its advocates, however, still held that these abilities formed a systematic scheme, not a miscellaneous aggregate: "intellectual qualities," says Binet, "are not simply superposable: on the contrary, there is a classification, a *hierarchy* among the diverse intelligences" (3, p. 40).

Statistical Corollaries. The statistical consequences of these alternative conceptions were fairly obvious. Galton, Pearson, and their followers argued that, in all individuals belonging to the same biological species, even 'unlike characters' would show a positive if somewhat low degree of correlation, while 'like characters' would yield much higher correlations.³ In keeping with this doctrine, Cattell and Wissler (who first suggested combining the Galton-Pearson methods with the testing techniques of laboratory psychology) believed that mental activities would also reveal a distinction between what they called 'general' and 'special' abilities.

Spearman took the opposite view. Accepting from Spencer⁴ and Ward⁵ the 'theorem of intellectual unity,' he argued that all cognitive processes were merely differentiations of a single general function: their apparent diversity arose solely from differences in mental content (sensations, images, words, school subjects, or the like), not from differences "in the actual form of mental activity," as *die alte Vermögenstheorie* had erroneously supposed. The 'hierarchy of specific intelligences'⁶ must therefore depend on a purely quantitative variation, and must consist in a continuous transition from simple processes to more complex rather than in a steplike series of 'levels' different in quality or kind. His view of

¹ Sully (Spearman's predecessor at University College) seems first to have used the term 'factors' to designate the principles of classification in psychology (*The Human Mind*, 1892, I, p. 70): he accepts the theory of 'levels' (p. 74), and reminds us that "the view of mind as a hierarchy of graded powers" goes back to "the elaborate scheme of Plato" (II, p. 327). Ward's advocacy of a monistic view was so strong that he even objected to the very idea of 'classification' in psychology: see Sully's reply, *loc. cit.*, p. 71. Bain attempts to classify cognitive faculties into '*genera* and *species*' in *The Senses and the Intellect*, App. I on 'Classifications of Intellectual Powers'. American psychologists, influenced largely by James and later by Thorndike, have always tended to favour a strongly pluralistic interpretation: according to this view, as Binet aptly puts it, "l'esprit ne serait qu'une collection absolument *hétéroclite* de facultés qui restent rigoureusement indépendantes et sont comme juxtaposées."

² Binet uses the phrase 'general intelligence' when he is emphasizing the contrast between an all-round intellectual level and limited 'aptitudes' like 'musical memory' or mechanical ability (e.g., 'The Development of Intelligence,' p. 39); he uses the phrase 'natural intelligence' when he is emphasizing that the general ability he proposes to measure is the effect of innate constitution, not of 'instruction' at school or during post-natal experience generally (e.g., *ibid.*, p. 42).

³ *Grammar of Science* (1900), pp. 403f. Elderton distinguishes three levels of correlation in human data (*Primer of Statistics*, 1909, Table XIV, p. 70).

⁴ "Intelligence neither has distinct grades, nor is it constituted of faculties that are truly independent" (Spencer, H., *Principles of Psychology*, 1872, I, p. 388).

⁵ Ward, J., *Encyclopædia Britannica*, s.v. 'Psychology,' p. 74.

⁶ It should be noted that the term 'hierarchy' was applied by Spearman in the same way as other psychologists had applied it, namely, in reference to the mental abilities or processes themselves, not to a table of correlation coefficients: (cf. *Am. J. Psych.*, IV, pp. 274, 284). But, as McDougall and others pointed out, "Spearman's oversimplified insistence on the *order* of intellectual abilities seemed a retrograde step to the older conception of *linear* gradation—the very notion which the word 'hierarchy' was intended to correct."

the mind was therefore (as he expressed it) strictly 'unifocal' (6, p. 53). It followed that, instead of two sharply distinguishable grades of figures—low coefficients where only 'general ability' was involved, and clusters of high coefficients where 'special abilities' were super-added—the correlations between different cognitive activities should display even and uniform gradations, such as Galton had described where only a single causal tendency was at work.¹

Spearman's first experiments, unlike those of Wissler, were confined to two or three tests of one mental level only, namely, sensory discrimination. This choice he defended on the grounds that, as Bain and Sully had maintained, discrimination was the essential element in all cognitive processes, and that discrimination alone could be accurately measured by laboratory tests. But correlations based on such limited data could hardly be expected to afford an adequate support for his somewhat sweeping criticisms of those contemporary investigators who had already tried Pearsonian methods, like Wissler and Thorndike, or of earlier writers, like Galton and Binet, who still adhered to the doctrines of 'faculties' and 'types.'

My own early efforts were prompted by the notion that the 'pluralistic' doctrines of the older faculty psychologists and the 'monistic' trend in contemporary British psychologists were not, in point of fact incompatible, except when stated in a needlessly uncompromising shape. After all, both Galton and Binet had, vaguely and implicitly no doubt, combined the two views; and quite recently Meumann had expressly proposed a 'modified psychological classification' which should include both *allgemeine* and *spezielle oder qualitativ bestimmte Fähigkeiten*, and even sketched a programme of tests on this eclectic basis.² But in order to establish a composite hypothesis of this kind, it seemed necessary to cover the widest possible variety of mental processes, including in the long run all the main levels and aspects of the mind, 'higher mental processes' as well as lower, emotional or conative processes as well as intellectual or cognitive (5, 7, 9).

The Hierarchical Organization of Mind. With this extended scheme, unmistakable indications emerged, almost from the start, of something like a systematic *structure*³ in the mind; and, as the investigations spread over an ever-increasing range, the general outlines of this 'structure' became more and more apparent. The mind, in fact, appeared to be progressively organized into a system of factors of varying degrees of generality, the more general factors including the more specialized, as countries include counties, and counties include parishes and towns. This, as I ventured to point out, was fully in keeping with the original scholastic view of mental

¹ In his later articles Spearman endeavoured to strengthen his monistic standpoint by invoking the same concept that had done so much to unify physical science in the 19th century. In his joint paper with Hart he argued that every intellectual activity could be regarded as a manifestation of 'mental energy' (6). The hierarchical arrangement of different mental activities was therefore due simply to the fact that each was 'saturated' with this common energy to a different degree. This offered a still more plausible ground for expecting a 'uniform or even gradation' in the relations between different mental activities (10, p. 137).

² Meumann, E. *Experimentelle Pädagogik*, (1907) I, pp. 73f.; also his earlier articles on *Intelligenzprüfungen an Kindern*, summarized pp. 385f. of his book. As was mentioned in my own paper, several of my 'tests of higher mental functions' were developed out of Meumann's suggestions. As a guiding hypothesis I took the "hierarchical scheme of four cognitive levels" (5, pp. 98-9) adapted by McDougall from Sully (*Outline*, pp. 43-4), and perhaps most clearly formulated in the well-known textbook by Mellone and Margaret Drummond (*Elements of Psychology* 1907, pp. 71f).

³ This term, like many others used in discussions of that day, was also taken from McDougall, to whom my own work owed so much. To McDougall, too, is due the most explicit formulation of what was for long the central problem in this field of work: "the question" (as he puts it) "whether knowing is the exercise of a *single* faculty, or whether we must recognize a *variety* of modes of knowing, each the exercise of a distinct faculty" (*Psychology: The Study of Behaviour*, 1912, chapter III, on 'The Structure of the Mind,' p. 79). At that date his own tentative answer was that, in addition to the "potentiality of knowing or thinking in general," we ought also to recognize a general faculty of striving and another of feeling, and that each of these should be envisaged, not as single self-sufficient functions, but as covering "a *class* of faculties of similar nature. . . . So much we have to concede to the theory of faculties, and take over from it" (*loc. cit.*, pp. 78-9: see also the quotation from his later book overleaf).

faculties, which regarded them as classifiable into subaltern *genera* and *species*.¹ McDougall himself strongly favoured some such interpretation of the facts. On the cognitive side, as he wrote in a later work, "examination of the results of correlation leads us to a view that the multitude of abilities are organized in many groups; these in turn in systems of allied groups; and these again in wider systems: all the systems having been formed by the growth and differentiation of innate abilities."²

Spearman replied by reprinting (6, p. 54) the first of my own correlation tables, and proving, or seeking to prove, that, if sampling errors were borne in mind, it supplied little or no evidence of discrete 'levels' or of groups of magnified coefficients.³ However, to-day, I fancy, most psychologists familiar with factorial work would feel far more assured of the so-called 'group factors' than of those more general factors, which serve to co-ordinate activities between the different 'groups'. And, whichever side they are disposed to take in the Spearman-Pearson controversy, the majority would probably endorse Prof. Godfrey Thomson's view that "the mind is comparatively structureless."⁴

So long as the number of persons tested was small, the sampling errors remained so high that even the most sanguine investigator could scarcely hope to distinguish more than one or two main factors in a single research; and the way in which the different factors might be related or arranged remained for the most part a matter for conjecture. But, now that data from far larger samples have become available, it seems desirable to consider whether the older and simpler techniques are really competent to elicit or establish a systematic factorial structure, such as earlier speculations suggested.

Examples. Even with the crude methods adopted in the older enquiries, a subdivision of broader factors into narrower was occasionally discernible. Consider, for example, the results of the first attempts at testing the educational abilities of school children: in addition to the ubiquitous 'general factor,' there was ample evidence for a broad group factor

¹ Cf., for example, Thomas Aquinas, *Summa Theologica*, Quaestiones, LXXVII, LXXVIII. To avoid misconception let me recall that in scholastic logic the terms *genus* and *species* were correlative, not fixed: this usage is a little obscured by their place in modern scientific classifications, where they appear as the last divisions of the familiar series, 'kingdom,' 'class,' 'order,' 'genus,' and 'species.' Modern writers on taxonomy seem to suppose that the idea of 'hierarchical classification'—a *systema naturae* in place of the old linear *scala naturae*—did not arise until about the time of Linnaeus and the early evolutionists (cf., for instance, Högben, L., *Science for the Citizen*, p. 928). But the scholastic philosophers, following post-Aristotelian logicians, made full use of it in their speculations; and their faculties formed a system, not a mere heterogeneous list.

² McDougall, *Energies of Men*, 1932, p. 95. As regards Spearman's two-factor theory he observes: "very little reflection shows that such a view is too simple to interpret the facts of correlation"; and he considers that, at the time he was writing, "Spearman himself, as well as his various critics, seems to be moving in the direction of the [alternative] view," as summarized in the statements quoted from McDougall above.

³ *Loc. cit.*, p. 54. Spearman's imaginary scheme of the 'high' and 'low' coefficients to be expected on the multifactor hypothesis advocated by his opponents (*ibid.*, table on p. 57) provides the clearest picture of the issue between us. It serves admirably to elucidate Pearson's original conclusion. Pearson himself, of course, was referring to biological characteristics in general; but his followers including Thorndike, Wissler, and Brown, had unanimously assumed that it would also hold good of psychological. And it was this that Spearman so stoutly contested.

⁴ *The Factorial Analysis of Human Ability* (1946), p. 312. In America most factorists appear to follow Thurstone in assuming that the mind may be regarded as consisting of a heterogeneous collection of 'primary abilities.' Like atoms in chemistry, they are described as limited in number (not much more than a dozen apparently), and as being 'fundamental' or 'simple,' i.e., indivisible. My view of mental structure is the exact opposite. It might be expressed in Russell's words: "The analysis of structure proceeds by successive stages. The ultimate units so far reached may at any moment turn out to be capable of further analysis" (*Human Knowledge*, p. 268). The number of 'abilities' enumerated is thus limited only by knowledge and convenience. It is this progressive subdivision of factors that gives rise to a hierarchical system.

concerned with verbal abilities of every kind ; but this seemed to split into two sub-factors concerned with two distinguishable types of verbal operation—viz., the use of words in isolation, and the use of words in a context.¹ Now most investigators who have observed the presence of two or more verbal factors have treated them as constituting two disconnected group factors, each subsisting independently within its own narrow field. But, on scrutinizing the actual figures, it will usually be seen that the pattern of correlations obtained after the general factor has been eliminated strongly suggests a third and broader verbal factor covering or including both these more specialized factors—a factor for the *genus* embracing, as it were, sub-factors for the two main *species*. Other factors, e.g., the 'numerical' or 'mathematical factors' and the 'logical' or 'reasoning factors,' seem also to contain sub-factors within themselves, arranged in much the same way.

A still more striking example is to be found in recent work on æsthetic appreciation. Here we commonly discover, first, a relatively general factor entering into æsthetic appreciation of every kind ; this, however, appears to include two or more moderately specialized forms—e.g., factors for the appreciation of visible art-forms and of audible art-forms ; and these in turn (when the tests employed permit it) become further subdivided into factors for the appreciation of colour and of form in the first case and of melody and harmony in the second case.²

One of the most elaborate illustrations of this type of factorial structure appears in the last issue of this journal. In Dr. Banks' re-analysis of Prof. Cattell's large table of personality traits, the complete scheme indicated by the data would contain eight narrow group factors (each containing three to five traits only) bracketed in pairs under four broader factors (each containing eight or nine traits), while these in turn would appear to be subsumed under two still wider factors, covering between them the whole of the table.³

Inadequacy of Existing Factorial Procedures. The ordinary methods of factor analysis do not envisage any systematic scheme or hierarchical differentiation of the kind I have described. On the contrary they are apt to obscure it. Certainly with 'simple summation' and the 'centroid' method, as ordinarily carried out, the first factor is usually a general factor, common to all the traits ; the second then subdivides them into two broad classes, positive and negative ; and the third factor usually subdivides each of these into two subclasses. But in this last stage the same principle of division is run throughout the entire list of traits or tests. This is as though we were to insist that, because in analysing musical appreciation we encounter a contrast between melody and harmony, therefore in pictorial appreciation the same contrast between melody and harmony must reappear.

The difference between the two principles can perhaps be exhibited most simply by comparing the genealogical trees or *stemmata* which logicians have used to illustrate what they term 'cross classification' and 'divergent classification' respectively (Figures 1 and 2). In the ordinary factorial procedure two bipolar factors necessarily imply a complete 'cross classification'—a type of classification which, in point of fact, is rather rare with empirical material. The procedure which I am now suggesting is designed to allow of 'divergent classification'.

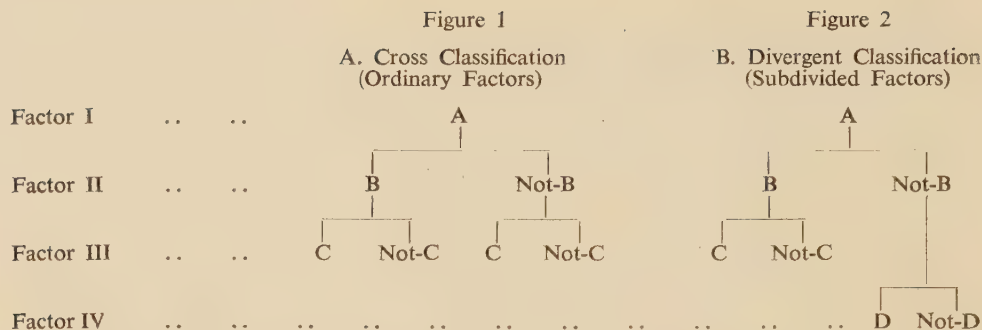
A stock example of the first type of classification is the Aristotelian division of (A) all material substances into (B) solid or 'dry' and (*not-B*) liquid or 'wet' ; each of these is then subdivided into (C) 'warm' and (*not-C*) 'cold'.⁴ As an instance of the second type we may take G. Bentham's

¹ Burt, C. (1917), *Distribution and Relations of Educational Abilities*, p. 59. Later work has suggested yet further subdivisions of the broad 'verbal factor' ; and here, as in so many other instances, the mental structure revealed by factor analysis corresponds very closely with the 'neurological architecture' tentatively described by those who have approached the problem from a pathological or clinical standpoint (e.g., Head in his classification of the aphasias).

² Cf. Williams, Winter, and Woods, *Brit. J. Educ. Psych.*, VIII, pp. 265-84, and articles and theses by Dewar, Eysenck, and Wing on æsthetic appreciation.

³ This *Journal*, I, p. 209f.

⁴ *De Corr. et Gen.*, II, 3. Note that the bipolar nature of the classification is indicated by the fact that (as Aristotle himself points out) two out of the six possible combinations (hot with cold, wet with dry) yield null classes. One of the oldest and most elaborate instances of cross-classification of which I am aware is that of 'feet' in Graeco-Latin prosody (Quintilian, IX, iv, 81) : with some simplification the traditional names would furnish convenient terms for describing the sign-patterns of traits.



division¹ of (*A*) living organisms into (*B*) those possessing locomotion (animals) and (*not-B*) those not possessing locomotion (plants); the former are then subdivided into (*C*) animals possessing backbones (vertebrates) and (*not-C*) animals not possessing backbones (invertebrates); the latter, however, are subdivided into (*D*) plants producing seeds (seminates) and (*not-D*) plants not producing seeds (inseminates). At the second stage in such classifications, one and the same bipolar factor would serve to subdivide *both B and not-B* only in the first example, not in the second. Clearly, if a positive saturation in the second bipolar is taken to mean 'seed-producing', a negative saturation cannot be used to mean 'devoid of a backbone'.²

It might be supposed that we could circumvent these difficulties by rotating the bipolar factors to those of a 'simple structure.' But a 'simple structure,' if I understand the conception rightly, consists simply of a number of irregularly overlapping group factors. There is no comprehensive general factor; nor do the broader group factors systematically include the narrower. Nowhere in the descriptions given of it, or in the examples so far published, do we find that the 'structure' is systematically organized.

For a psychological illustration of the main differences we may briefly glance at the example which Thurstone employs to explain both his method of 'centroid' analysis and his method of rotating the centroid factors to procure a 'simple structure.' In Table 2 (11, p. 167) he gives the loadings for the three main centroid factors³—the only factors "large enough to justify serious consideration." The first is the usual 'general factor' common to the 15 tests. The two bipolar factors then introduce a cross-classification, dividing the entire battery of tests into four main sets. Factor II evidently classifies them into verbal and non-verbal. Factor III perhaps might be considered as cross-classifying them into routine tests and logical tests. Actually it separates the 6 verbal tests into (*a*) 3 more elementary 'opposites' tests and (*b*) 3 more complex logical tests (definitions, analogies, etc.); and at the same time it subdivides the 9 non-verbal tests into (*a*) 5 spatial tests and (*b*) 4 numerical tests. But are the two modes of subdivision really the same in principle? There is a simple test. If this double subdivision is to be effected by one and the same factor, then there should obviously be a set of significant cross-correlations linking the spatial tests with the 'opposites' tests, and the numerical tests with the logical tests. But, as we shall see in a moment, the evidence for any such cross-correlations is exceedingly doubtful.

¹ *Essai sur la classification* (1823). Cf. 14, pp. 108f.

² From the logical standpoint the best examination of the general problem is (so far as my acquaintance goes) that contained in Sigwart's chapter on 'Systematization in Classificatory Form' (*Logic*, II, 1893, pp. 519f.); the need, and even the general idea, of factor analysis in anthropometric material is strikingly anticipated in his paragraph on how to classify human skulls (p. 527). Venn states that "the object of all Classification" is to arrange things in a "hierarchy of successive classes" by a process of "continued subdivision." His discussion brings out the difficulties of achieving 'economy' ("maximum amount of assertion with a minimum number of propositions") by the 'numerical' procedures known in his day (*Empirical Logic*, 1889, pp. 322, 332-3, 83-4).

³ Reproduced from his Table 25 (11, p. 117). Referring presumably to my own earlier endeavours to interpret bipolar factors, Thurstone observes (p. 113) that "it is an error, frequently made, to attempt a psychological interpretation of the factors which have not been rotated," such as those in Table 25 or 2. I agree that errors may be committed in such interpretations; that, indeed, is my argument here. But I do not believe that the attempt itself is necessarily an error.

Thurstone himself, of course, would rejoin that such difficulties only arise when we endeavour to read a meaning into the preliminary centroid analysis : to discover the concrete nature of the factors we must first carry out a rotation. After rotating the centroid factors to reach a 'simple structure,' he himself finds only three group factors (Table 4, p. 169) : (A) a 'number factor,' (B) a 'verbal factor,' and (C) a factor 'concerned with visual imagery and perhaps kinæsthesia'. My fourth group of tests, consisting of the three harder verbal tests, are treated as composite : with these half of factor A is made to overlap with half of factor B. And Thurstone tentatively suggests that this is because what he has provisionally called a 'number factor' may be concerned, not merely with "number as such," but with "some kind of facility for logical thinking" required for solving mathematical problems, but not peculiar to it ; in short (though Thurstone himself would probably not accept this inference) that it is rather like the general factor of 'intelligence,' as interpreted in this country. However, a more plausible pattern could be obtained if we made the broader verbal factor divisible into two sub-factors, and kept the spatial and the number factors distinct.

In my own view Thurstone's broad factor, labelled A, entering into half the verbal tests and half the non-verbal, is an artefact. The misleading notion that there must be some such factor is suggested by the similarity of signs in the second bipolar (factor III), where it extends over both the verbal and the non-verbal groups. As will be seen from Table 2 (p. 167), this factor divides both the positive and the negative sections of factor II into a positive and a negative subsection, and then seems to treat its own two positive subsections as parts of one and the same group factor. But, as I have already contended, were this valid, we should expect to find significant residual correlations between these two sets of tests, viz., (a) tests 1, 3 and 4, and (b) tests 6, 7, 8 and 9. On turning back to Table 17 (p. 112), however, the largest residual (for tests 1 and 8) is only .071. The probable error of the observed correlation is $\pm .030$: so that even this residual is devoid of statistical significance. Further, its effect is largely neutralized by the negative residual correlation between 1 and 9. At this stage of the analysis my own preference would be to substitute two independent bipolar factors of narrow range, one covering only the verbal tests and contrasting tests 1, 3 and 4 with 2, 5 and 10 (but not with 6, 7, 8 and 9), the other covering only the non-verbal, and contrasting tests 11, 14, 15, 17 and 18 with 6, 7, 8 and 9.¹

Criterion. If this hierarchical type of factor matrix is to be proposed as an alternative to the more familiar pattern, it will not, in my opinion, be sufficient to let it emerge as a result of repeated rotations. That would in most instances leave the effective decision to conscious or unconscious preconceptions. We have seen, however, that there is a simple and objective criterion for verifying the kind of structure I have described. If the two narrow bipolars are independent, then, after the general factor and first bipolar have been eliminated, the residual cross-correlations between the tests forming the positive section of factor II (verbal tests) and those forming its negative section (non-verbal tests) should be approximately zero. In terms of Figure 2 (p. 46), there can be no correlation of *C* with *D*, or of *not-C* with *not-D*.

On examining Thurstone's own table of residuals (p. 112), we find that out of the 54 cross-correlations 16 are less than 0.01 ; only four are greater than 0.06 (i.e., greater than twice the probable error), and these occur in two tests only ; while none approach three times the probable error. On the other hand, out of the 51 inter-correlations (on which alone the narrow bipolars would rest) only two are less than 0.01, and as many as 10 are greater than 0.06.

It would seem, then, as I have already stated elsewhere, that "the use of a more elaborate statistical procedure demonstrates (what was suspected from the start) that most of the group factors previously recognized are not single self-contained abilities, but unfold into factors still more specialized. 'Verbal ability,' for instance, proves to be highly complex—the ability for understanding isolated words being relatively independent of facility in understanding verbal patterns ; similarly, the 'spatial factor' may be visual or kinæsthetic ; 'comprehension' may involve either visual imagery or verbal," and so on (Burt and John, 1942, p. 161). Accordingly,

¹ There is some evidence that the 'number factor' is also divisible into two sub-factors, not unlike those required for 'problem arithmetic' and 'mechanical arithmetic' respectively. By itself, this one correlation table, based on only 356 examinees, is not sufficient to support a pattern comprising so many factors. But the scheme I have suggested seems borne out by other tables in Brigham's original collection. However, my object at the moment is merely to illustrate the modified scheme by a readily accessible table, not to demonstrate it.

in any given enquiry that is sufficiently extensive, we tend to get a pyramid of factors, with the primary group factor—the all-inclusive ‘general’ factor—at the summit, and secondary, tertiary, and yet other group factors ranged in succession below: probably the clearest way to represent these ramifying patterns is to arrange the factors in the form of a genealogical tree, like those depicted in the article just cited or in the diagrams on page 46 (cf. Burt and John, *loc cit.*, Fig. 1, p. 125). In short, what all these analyses irresistibly suggest is a classification of mental attributes into a hierarchy in which the *summun genus* is divided into two or three broad *genera intermedia*, each of these into narrower *sub-genera*, these in turn into *species*, and the species into *sub-species*, narrower still, rather like the progressive schemes with which we are familiar in the classification of animals or plants.¹

II. DEMONSTRATION FROM BODY-MEASUREMENTS: DATA

Problem. So far I have been concerned chiefly to explain the origin and nature of the hypothetical scheme that I am advocating. It is now incumbent on me to give at least one clear-cut case in which the hypothesis can be conclusively verified, and to illustrate in fuller detail the working procedure proposed.

The need for some modified method of factor analysis, such as I have just suggested, can best be demonstrated in the case of bodily measurements. Physical data have the special advantage that the errors of measurement are themselves far smaller than in the case of psychological traits; consequently emphasis can more safely be laid on minor divergences from correlational patterns of the stock kind usually assumed. Moreover, we have here a good deal more supplementary evidence, of a non-statistical kind, to corroborate our statistical inferences about the types so revealed.

In previous articles I have reported the results of several tentative enquiries into the nature of body-types and temperamental-types respectively (7, 19, 20, 21, and refs.). The primary object of the investigation now to be described was to examine the correlations between the two. Earlier instalments of the data have already been published elsewhere (12; cf. 9, p. 15, 14, p. 387 and refs.).

The list of physical measurements selected for this and similar enquiries was based originally on the scheme proposed by Viola.² It was varied or extended to include a few

¹ A number of similar ‘pedigree-tables’ of factors will be found in Banks (16) figures facing pp. 6, 23, 104, and 187 (data from tests applied in the Army and Navy). With such diagrams in mind, and partly perhaps to avoid the term ‘hierarchy’ (which in factor analysis has become diverted to express a very different concept) Cattell has suggested using the name ‘genealogical’ “to describe the relationships . . . in Burt’s system” (18, p. 280, footnote 1): the word has the further advantage of suggesting that the relations may be genetic—the narrower factors being differentiated out of the broader. However, now that most psychologists have become familiar with the phrase ‘matrix of rank one,’ the term ‘hierarchy’, I suggest, might be restored to its former use (a use already revived by non-statistical writers like Nunn and Gesell). Strictly, as its etymology indicates, it implies that the more general activities originate or govern (*ἀρχή*) the more specialized. But for the factorist, as for the logician, it must be taken as describing in the first instance merely “a method of classification by a system of terms of successive rank, e.g., ‘a h. of concepts’, Bowen, *Logic*, 1864, ch. vi” (*Oxford Dict.*, s.v.). For a more philosophical definition see Whitehead, A. N., *Science and the Modern World*, pp. 241-2; and for a definition of its meaning as applied to an individual organism, see J. H. Woodger, *Biological Principles*, pp. 311-2. My own definition is given on p. 63.

² *La tecnica antropometrica a scopo clinico*, Milano, 1905. For a conveniently accessible summary, see Naccarati, S., ‘The Morphologic Aspect of Intelligence,’ *Arch. Psych.*, XLV, 1921, pp. 4f. and Figure 1.

TABLE I. CORRELATIONS BETWEEN PHYSICAL TRAITS

Traits	1	2	3	4	5	6	7	8	9	10	11	12
1. Upper Arm Length ..	(.753)	.639	.737	.745	.475	.412	.603	.536	.512	.225	.214	.151
2. Lower Arm Length ..	.639	(.657)	.656	.768	.366	.397	.529	.423	.371	.183	.112	.092
3. Upper Leg Length ..	.737	.656	(.794)	.804	.488	.439	.616	.555	.538	.247	.176	.143
4. Lower Leg Length ..	.745	.768	.804	(.988)	.510	.571	.536	.422	.356	.174	.073	.106
Sum ..	2.874	2.720	2.991	3.305	1.839	1.819	2.284	1.936	1.777	.829	.575	.492
5. Sitting Height ..	.475	.366	.488	.510	(.785)	.723	.385	.458	.344	.129	.134	.110
6. Sternal Length ..	.412	.397	.439	.571	.723	(.753)	.363	.387	.285	.150	.113	.108
Sum ..	.887	.763	.927	1.081	1.508	1.476	.748	.845	.629	.279	.247	.218
7. Shoulder Breadth ..	.603	.529	.616	.536	.385	.363	(.909)	.761	.805	.416	.319	.179
8. Chest Breadth ..	.536	.423	.555	.422	.458	.387	.761	(.709)	.738	.365	.323	.184
9. Hip Breadth ..	.512	.371	.538	.356	.344	.285	.805	.738	(.911)	.547	.532	.338
Sum ..	1.651	1.323	1.709	1.314	1.187	1.035	2.475	2.208	2.454	1.328	1.174	.701
10. Chest : Ant-post. Diam. ..	.225	.183	.247	.174	.129	.150	.416	.365	.547	(.442)	.458	.321
11. Abdomen : Ant-post. Diam. ..	.214	.112	.176	.073	.134	.113	.319	.323	.532	.458	(.598)	.389
12. Subcutaneous Fat ..	.151	.092	.143	.106	.110	.108	.179	.184	.338	.321	.389	(.287)
Sum ..	.590	.387	.566	.353	.373	.371	.914	.872	1.417	1.221	1.445	.997
Grand Total ..	6.002	5.193	6.193	6.053	4.907	4.701	6.421	5.861	6.277	3.657	3.441	2.408

TABLE II. FACTOR SATURATIONS BY SIMPLE SUMMATION

Traits	A. Ordinary Factors					B. Subdivided Factors				
	I	II	III	IV	V	I	II	III	IV	V
1. Upper Arm Length	.768	.248	.199	.120	.174	.768	.248	.230	—	.147
2. Lower Arm Length	.664	.325	.225	.169	.126	.664	.325	.281	—	.143
3. Upper Leg Length	.792	.274	.221	.105	.118	.792	.274	.238	—	.100
4. Lower Leg Length	.774	.535	.136	.222	.166	.774	.535	.245	—	.220
Sum ..	2.998	1.382	.781	.616	.000	2.998	1.382	.994	—	.116
5. Sitting Height	.628	.334	-.374	-.345	.145	.628	.334	-.507	—	.205
6. Sternal Length	.601	.362	-.407	-.271	-.145	.601	.362	-.487	—	-.089
Sum ..	1.229	.696	-.781	-.616	.000	1.229	.696	-.994	—	.116
7. Shoulder Breadth..	.821	-.212	.329	-.232	-.136	.821	-.212	—	.396	-.073
8. Chest Breadth	.750	.187	.153	-.281	.069	.750	.187	—	.308	.137
9. Hip Breadth	.803	-.476	.115	-.103	.067	.803	-.476	—	.154	.081
Sum ..	2.374	-.875	.597	-.616	.000	2.374	-.875	—	.858	.145
10. Chest : Ant-post. Diam.	.468	-.418	.104	.140	-.104	.468	-.418	—	-.173	-.136
11. Abdomen : Ant-post. Diam.	.440	-.500	-.263	.245	.080	.440	-.500	—	-.359	.029
12. Subcutaneous Fat..	.308	-.285	-.230	.231	.024	.308	-.285	—	-.326	-.038
Sum ..	1.216	-1.203	-.597	.616	.000	1.216	-1.203	—	-.858	-.145
Sum of Squares	5.402	1.584	.743	.572	.175	5.402	1.584	—	.541	.109
Contribution to Variance	45.0%	13.2%	6.2%	4.8%	1.5%	45.0%	13.2%	—	4.5%	0.9%

additional measurements, chiefly suggested in discussions with my colleagues, Dr. F. C. Shrubbsall (at that time Secretary of the Anthropological Section of the British Association and Convener of the Measurements Subcommittee) and Dr. James Kerr (then Consulting Medical Officer for London). It was further modified from time to time, as less reliable or less suggestive measurements were discarded, and others introduced. For the most part the methods of taking the measurements followed the instructions laid down by the British Association Committee on Anthropometrical Investigations.¹

Our earlier attempts at factorizing the measurements thus obtained indicated that the resulting factors may often be obscured by including traits (like height or weight) which overlap widely with other traits or (like the circumferential measurements) depend on more than one linear dimension. Accordingly, in the following analysis height and weight have been omitted, and measurements of breadth and thickness have been retained in preference to circumferential measurements. The resulting list provides a fairly well-balanced set; but it is perhaps regrettable that only two measurements are available for length of trunk.

The following are the traits finally retained for purposes of the present analysis :

(1) Length of right upper arm (measured from the outer margin of acromion to the lowest point of the external condyle of the humerus) ;

(2) Length of right lower arm (i.e., forearm, measured from lowest point of the external condyle to the middle finger tip) ;

(3) Length of right upper leg (i.e., thigh measured from upper edge of great trochanter to margin of superior extremity of tibia on outer side of knee joint) ;

(4) Length of right lower leg (measured from extremity of tibia to sole) ;

(5) Sitting height (length of back, measured by Dreyer's method ²) ;

(6) Length of sternum (length of front of chest) ;

(7) Breadth of shoulders (measured with callipers between outer margins of right and left acromion) ;

(8) Breadth of chest (transverse thoracic diameter measured, at level of junction of fourth rib-cartilage with sternum, during complete expiration) ;

(9) Breadth of hips (maximum transverse pelvic diameter, measured with callipers between the iliac crests) ;

(10) Thickness of chest (antero-posterior thoracic diameter at same level as 8) ;

(11) Thickness of abdomen (antero-posterior abdominal diameter at same level as 9) ;

(12) Thickness of subcutaneous fat (average of four measurements).³

Subjects. The majority of the measurements were obtained for children of every age during the elementary school period. Here I propose to confine myself to the two oldest and largest age-groups, namely, boys aged 12 and 13 last birthday. These numbered 457 in all. The last three measurements, however, were made on part of the sample only, namely, 393.

Correlations. The correlations for each age-group were factorized separately. The age-groups with which we are here concerned revealed very little difference. Accordingly it seemed legitimate to pool the figures for these two groups. The average correlations thus obtained between the twelve traits are shown in Table I.

¹ 'Report of Subcommittee on Anatomical Measurements,' 1908, pp. 359f. Cf. also Kerr, J., *Fundamentals of School Health* (1926), pp. 13-52. and refs., and, for a brief discussion of the problem of body-types in children, *ibid.*, pp. 76f. and refs. I have to acknowledge my indebtedness to those who co-operated in collecting the measurements. In the earlier part of the work I enjoyed the help of Dr. Lucy Hoesch-Ernst. Later we were assisted by teachers and school medical officers who developed an interest in the project. Unfortunately, after the first few groups had been measured, it was decided to drop the transverse and the antero-posterior (sagittal) measurements (which furnished rather low reliability coefficients) and substitute circumferential measurements. It was only as the analysis progressed that the importance of the former became evident, and a further series was obtained in which the transverse measurements and the measurements of subcutaneous fat were reintroduced.

² See Kerr, *loc. cit.*, p. 15.

³ Estimated by pinching a vertical fold of skin at specified points over the ribs, abdomen, right upper arm, and right thigh, and measuring its thickness with callipers : cf. Kerr, *loc. cit.*, pp. 115 and refs.

TABLE III. RESIDUALS AFTER REMOVING THE GENERAL AND FIRST BIPOLAR FACTOR

Traits	1	2	3	4	5	6	7	8	9	10	11	12
1. Upper Arm Length ..	.102	-.049	.061	.018	-.090	-.139	.025	.007	.013	-.030	.000	.015
2. Lower Arm Length ..	.049	.110	.041	.080	-.159	-.120	.052	-.014	-.008	.008	-.018	-.020
3. Upper Leg Length ..	.061	.041	.092	.044	-.101	-.137	.023	.012	.032	-.009	-.036	-.023
4. Lower Leg Length ..	.018	.080	.044	.102	-.155	-.089	.013	-.058	-.011	.036	.000	.020
5. Sitting Height ..	-.090	-.159	-.101	-.155	.280	.225	-.059	.049	-.001	-.025	.024	.012
6. Sternal Length ..	-.139	-.120	-.137	-.089	.225	.260	-.054	.004	-.025	.020	.029	.026
7. Shoulder Breadth ..	.025	.052	.023	.013	-.059	-.054	.190	.106	.045	-.057	-.149	-.134
8. Chest Breadth ..	.007	-.014	.012	-.058	.049	.004	.106	.112	.047	-.064	-.101	-.100
9. Hip Breadth ..	.013	-.008	.032	-.011	-.001	-.025	.045	.047	.040	-.028	-.059	-.045
10. Chest : Ant-post. Diam. ..	-.030	.008	-.009	.036	-.025	.020	-.057	-.064	-.028	.048	.043	.058
11. Abdomen : Ant-post. Diam. ..	.000	-.018	-.036	.000	.024	.029	.149	-.101	-.059	.043	.154	.110
12. Subcutaneous Fat ..	.015	-.020	-.023	.020	.012	.026	-.134	-.100	-.045	.058	.110	.111

Residuals that are more than twice the standard error of the initial correlations (as given in Table I) are printed in heavy type.

III. BIPOLAR ANALYSIS

(a) *Analysis by General Factors.* An ordinary factor analysis by simple summation was undertaken first of all. The requisite communalities were determined by successive approximation. This procedure is in my view essential. The calculations are not so lengthy as is commonly supposed. By adopting suitable short-cuts, they can easily be abridged. Working methods are illustrated in the appendix. In all, six factors were extracted. Only five appear to be capable of any concrete interpretation; and, judged by the usual criteria, only the first four are statistically significant.¹

The saturation coefficients for the first five factors are shown in Table IIA. It will be seen that this method necessarily makes every factor a 'general' factor: even the bipolars extend over the whole *totum divisum*.

The first accounts for 45 per cent. of the total variance. It is evidently a general factor for body size, with positive coefficients throughout, similar to that obtained in nearly all such analyses. The second accounts for 13 per cent. of the variance, and classifies the twelve traits into two main groups: (i) one group includes all the vertical or longitudinal measurements, (ii) the other all the horizontal or diametric measurements (with which fat-thickness may plausibly be classed). The third factor accounts for 6 per cent. of the variance: it subdivides (i) the longitudinal measurements into those relating to (a) length of limbs and (b) length of trunk respectively, and (ii) the horizontal measurements into (a) transverse or lateral and (b) antero-posterior or sagittal. The fourth factor accounts for 5 per cent. of the total variance—nearly as much as the third factor; and repeats the same classification with an altered sign-pattern. The fifth accounts for only $1\frac{1}{2}$ per cent. of the variance: and, with the present sample, can scarcely claim to be statistically significant. So far as it can be trusted, it seems to separate the measurements for the upper portions of the limbs from those for the lower portions of the limbs.²

Now, with this particular set of traits, the double use of the second bipolar factor, thus entailed by the ordinary factorial procedure, would appear to be quite indefensible on both *a priori* and *a posteriori* grounds. From the very nature of the traits involved it is illogical to suppose that the factor which contrasts lateral diameters with sagittal diameters should be the same as that which contrasts length of limbs with length of trunk. And empirically the employment of a single factor for this purpose could be justified only if the table of residuals on which it is based gave evidence of significant and systematic cross-correlations between the two former groups and the two latter: if, for instance, the limb-measurements correlated positively with the sagittal measurements, but negatively with the lateral measurements, and so on.

The residuals are shown in Table III. Figures that are statistically significant are printed in bolder type. It will be seen that the significant residuals are all confined to the correlations 'within groups'; while among the correlations 'between groups' not one even approaches significance.³ Including the diagonal entries, the residual inter-correlations between the first six traits average 0.101 (calculated

¹ Of the residuals on which the fourth factor is based, nine are numerically larger than $\pm .070$, and of these five are well over twice the standard error of the observed correlation. Of the residuals on which the fifth factor is based, only one is larger than $\pm .030$, and that is well under twice the standard error. I have to acknowledge the assistance of Miss G. Bruce in calculating the correlations and of Dr. C. Banks in calculating the factor-saturations.

² This factor (as we might expect from the 'centrifugal' tendency in growth) appears to correlate with an *infantile* body-build. The factor making for leg length at the expense of trunk length (equivalent to Meredith's 'skelic index') correlates with a *feminine* body-build.

³ Among the cross-correlations the only figures of any appreciable size (over 0.04 say) are apparently due to a slight overlapping of the growth of shoulder breadth and chest breadth with the long bone measurements. This is worth noting here, because it reappears in other surveys we have made, and is intelligible on anatomical grounds.

regardless of sign), and between the last six traits 0.076 ; whereas the residual cross-correlations of the first six with the last six average barely 0.012.

With the form of analysis we have used so far, however, the second bipolar factor necessarily assumes the existence of significant cross-correlations ; and it inevitably produces non-zero figures in the corresponding part of the hypothetical hierarchy that represents it. When in point of fact there are no such significant residuals, a third bipolar factor, with the same lines of division but a different sign-pattern, has consequently to be invoked to cancel the fictitious cross-correlations introduced by the second bipolar (see Table IIA).

(b) *Analysis by Subdivided Factors.* Such a compensatory procedure is as artificial and as clumsy as it is superfluous. There is, however, a simple way of surmounting the difficulty. We have merely to split the second bipolar into two, and then to limit each of the resulting bipolars to one set of traits alone, instead of requiring both to extend over every trait.¹ I propose to call factors so obtained 'subdivided factors,' because of their obvious analogy with the results of what is technically termed 'sub-division' in logic.²

Accordingly, after eliminating the general factor and the first bipolar, we now take the matrix of residuals, and factorize the two diagonal submatrices separately. The results of this alternative method of continuing the analysis are shown in Table IIB.

We see that the first of the limited bipolar factors divides the longitudinal measurements into two contrasted subgroups—those relating (a) to length of limbs and (b) to length of trunk ; the other subdivides the horizontal measurements into (a) lateral and (b) sagittal. The subdivisions are drawn at exactly the same points as before. But now, instead of one factor, we have two, viz., (i) a limb-versus-trunk factor, and (ii) a lateral-versus-sagittal factor ; and thus, as is logically required, the sub-classifications are effected by two principles of division instead of by one. These two limited factors³ account for 6.2 per cent. and 4.5 per cent. of the variance respectively. The fifth factor is but little altered, and remains non-significant.

IV. GROUP FACTOR ANALYSIS WITH SUBDIVIDED FACTORS

Alternative Physiological Explanations. In considering what may be the actual influences that determine body-build, the physiologist who seeks to identify abstract statistical factors, such as we have here obtained, with concrete causal factors will have

¹ This is equivalent to assuming that each of these limited factors will have zero saturations for those traits that belong to the complementary set : (see Table IIB, factors II and III). And this in turn is only warranted when every one of the residuals in the North-East and South-West quadrants is non-significant, as we have seen to be the case in the present instance. If, after all, a few of the residuals had turned out to be significant, we could take account of the overlap by using the formula for calculating the correlation of a given factor with traits extrinsic to the group on which that factor is based (13, p. 355, equations vii-ix).

² In logic 'division' is defined as 'the analysis of a more general class into its constituent sub-classes.' Both the term (*διαίρεσις*) and the technique were introduced by Plato, who seems to have thought of it as his chief contribution to methodology. It is interesting to note that his illustrations chiefly relate to the classification of individuals according to their habitual activities (*Soph.*, 221c-237a ; *Pol.*, 279b-305d). For the distinction between 'co-division' and 'sub-division', see Welton, J., *Logic*, I, p. 122.

³ Note that, for almost every trait, the saturation of the new limited bipolar is very nearly equal to the root of the square-sum of the two old general bipolars : e.g., $\sqrt{.199^2 + .120^2} = .230$ (approx.).

little difficulty in picturing appropriate agencies which could operate in a bipolar fashion. We can, for example (as certain writers have suggested), conceive of some endocrine secretion that might possess the property of increasing the length of the bones of the limbs, while diminishing the relative size of the bones of the trunk (as in certain forms of gigantism); and of other agencies which might tend to diminish the length of the long bones, while allowing the trunk to grow to a size that would be normal in an ordinary person, but looks disproportionately large in the patient (as in the achondroplastic dwarf). Nevertheless, the great majority of physiologists, I fancy, would contend that it is more in keeping with existing knowledge to postulate physiological agencies which have for the most part positive effects only, though varying in degree; so that contrasted effects will be produced by separate factors, not by the same. Accordingly, to conform with this second point of view, it may be helpful to factorize the observed table of correlations in terms of group factors whose significant saturations shall be exclusively positive. Which of the two kinds of factor-pattern is the more appropriate—the bipolar or group—is one of the many questions that have to be decided rather by external and material evidence than by a formal examination of the statistical results.¹

Non-Overlapping Factors. Let us now, therefore, transform our matrix of general and bipolar factors into a matrix of basic and group factors. As a rule, the best procedure is to start by constructing a provisional matrix of non-overlapping group factors, and then to undertake subsidiary calculations to allow for overlapping. This will usually lead to a circular process of progressive approximation.

The lines of division between the initial non-overlapping group factors will be determined by the preliminary bipolar analysis.² But, since we have found that the two complementary sections of the first bipolar factor subdivide separately into two distinct and limited bipolar factors of narrower range, we must also expect to find that the traits included in the first two group factors (each of which covers one section of the first bipolar) will similarly get subclassified into pairs of narrower group factors, corresponding to the subsections of the limited bipolars. In most of the published tables giving saturations for group factors, the pattern of significant figures forms, as it were, a single stair of block-like steps. Here we shall need a factor matrix showing at least two such stairs, a series of taller steps lying beneath and to the left of a series of shorter steps, which will have about half the height of the taller steps (see the figures in heavy type in Table IV).

¹ In theory a statistical test is certainly conceivable, and would evidently turn on the rank of the correlation matrix as disclosed by the preliminary bipolar analysis. For simplicity consider first the case in which a number of traits, forming a single *genus*, are divisible into two *species* but no further. If the two *species* can be exactly represented by one bipolar factor, the correlation matrix will have a rank of two only, since only two factors will be concerned—the general and the single bipolar. If, however, the two *species* are to be represented by two distinct *group* factors, it must (unless the group and basic factor saturations are proportional) have a rank of three; and in the bipolar analysis we shall get one general and two bipolars. The second bipolar will doubtless be comparatively small: nevertheless, it will be necessitated by the rank. Hence the theoretical test would consist in examining the residuals on which this third factor (the second bipolar) is based. In theory a similar increase in rank should also appear where we are concerned, not merely with two possible group factors, but with several subdivided group factors in addition.

In practice, however, such a test would be extremely difficult to apply. We could never expect an empirical correlation matrix to have a rank of exactly two or exactly three; and, unless the sets of traits were exceptionally well balanced and the sample of persons unusually large, it would be almost impossible to prove that the supplementary bipolar was an artificial consequence of analysing the product of a group factor matrix by the bipolar procedure.

² When the lines of division are automatically decided by the preliminary bipolar analysis, and the final rotation is based on an arithmetical calculation, then (apart from the minor changes introduced by further successive approximations) the results are uniquely determined within the conditions imposed. In my experience this is *not* true of 'simple structure' reached by a succession of graphical rotations. Certainly with the former method, the sign-pattern, and even the proportionate values of the saturation-coefficients, tend to be invariant from one investigation to another.

TABLE IV. GROUP FACTOR SATURATIONS

Trait	I	II	III	IV	V	VI	VII
1. Upper Arm Length ..	.758	.246	—015	.318	—011	—022	—039
2. Lower Arm Length ..	.557	.390	—017	.381	—039	—068	—021
3. Upper Leg Length ..	.735	.358	—056	.282	—013	—023	—043
4. Lower Leg Length ..	.562	.683	—014	.455	—024	—094	—036
5. Sitting Height ..	.539	.256	—051	—012	.703	—014	—077
6. Sternal Length ..	.341	.537	—074	—051	.572	—026	—043
7. Shoulder Breadth ..	.721	.207	.303	—029	—048	.459	—053
8. Chest Breadth ..	.682	—113	.262	—012	—114	.367	—022
9. Hip Breadth ..	.704	—025	.651	—044	—031	.230	—015
10. Chest : Ant-post. Diam.	.306	—042	.472	—033	—052	—094	.292
11. Abdomen : Ant-post. „	.273	—098	.488	—032	—008	—053	.471
12. Subcutaneous Fat ..	.167	—021	.332	—081	—024	—039	.403
Sum of Squares ..	3.826	1.230	1.167	.548	.843	.428	.484
Contribution to Variance ..	31.9%	10.3%	9.7%	4.6%	7.0%	3.6%	4.0%

Significant factor saturations are printed in heavy type.

So far as computation is concerned, this change introduces no new principle into the customary procedure for conducting a group factor analysis ; nor does it entail any novel formula. It merely means that the two 6×6 residual submatrices, left astride the leading diagonal after the basic factor has been removed, will themselves be subjected to a group factor analysis, instead of being treated as hierarchical matrices, each due to a single factor only. In the present case, on applying the usual formulæ and following the usual procedure, we obtain first a basic factor covering the whole set of traits ; then two group factors covering the first six and the last six traits respectively ; and finally four narrower group factors covering respectively the first four traits, the next two, the next three, and the last three. But with a simple pattern of non-overlapping factors of this kind, we shall get (as an experienced computer would see from the start) a number of rather large residuals suggesting that the first broad group factor (the factor entering into the first six traits) also overlaps on to two of the traits belonging to the complementary group factor, viz., breadth of shoulder and breadth of chest. However, this overlap can readily be allowed for by deriving the saturations for the first or basic factor from the last four traits only, instead of from the last six.

Overlapping Factors. If this is the only case of overlap allowed for in the matrix, and if a number of saturations of exactly zero are still retained, then we can hardly expect the seven group factors thus defined to fit the observed correlations as well as we could wish. To procure a factor matrix showing the same general scheme, but containing more precise and detailed figures, we must proceed to rotate the corresponding bipolar factors. The rotation is best effected arithmetically by the working-method described in previous papers. A hypothetical correlation matrix is first reconstructed from the set of non-overlapping group factors (here adjusted to allow of a special overlap) ; and this reconstructed correlation matrix is then factorized by simple summation. The transformation matrix needed to rotate the reconstructed bipolar factor matrix back to the original group factor matrix can be easily computed, and is of necessity perfectly orthogonal. This orthogonal transformation matrix is then applied to the actual matrix of general and bipolar factor saturations derived directly from the observed correlation matrix and already set out in Table IIA. Even so, the fit may be slightly improved by successive approximation.

The set of group factor saturations eventually reached is shown in Table IV. This yields as close a fit to the original table of correlations as is provided by the first five factors of Table II.

The group factors we have thus arrived at may be interpreted as follows :

I. The first is a basic factor for body size. The measurements showing the highest saturations with this factor are upper arm, thigh, shoulder breadth, and hip breadth : the sagittal measurements have decidedly low correlations with this factor.

This basic factor covers, or subdivides into, two broad group factors, viz.,

II A. A broad group factor for all length measurements. Here length of the lower leg yields the highest saturation : length of sternum yields the next highest. Owing to the overlap already mentioned, this factor exhibits a small but discernible correlation with shoulder breadth and chest breadth (possibly because the clavicle and ribs partake in part of the growth characteristic of the other ' long bones ').

II B. A broad group factor for all horizontal measurements. Hip breadth shows the highest saturation : with this sample the two sagittal measurements yield larger saturations than shoulder breadth.

Each of these broad group factors subdivides into two narrower group factors, viz.,

II A 1. A factor for length of limbs. Here the largest saturation is given by length of lower leg.

II A 2. A factor for length of trunk. The largest saturation is given by sitting height. (With this factor the large size of the saturations is apparently due in part to the fact that this group includes two traits only.)

II B 1. A factor for transverse measurements. The largest saturation is given by shoulder breadth.

II B 2. A factor for sagittal measurements, including thickness of subcutaneous fat. The largest saturation is given by abdominal depth.

The most important conclusions emerging from the foregoing analysis seem to be the following. First, the results fully confirm the view already suggested in previous papers, that the broader group factors revealed in earlier investigations often include, or split up into, smaller group factors, which can be conceived as more specialized differentiations of the broader factors under which they are subsumed. Secondly, and more particularly, the figures demonstrate beyond all question that the so-called ' pyknic ' type ¹ really includes two subtypes, viz., those who owe their pyknic character chiefly to the breadth of their bony skeleton, particularly of the thorax and pelvis, and those who owe it chiefly to the deposit of additional fat, particularly in the region of the abdomen.

The results reached by factorizing body-measurements for adults are quite in keeping with the conclusions here obtained for children. Thus, in an analysis of the correlations between 17 traits obtained from 528 R.A.F. recruits, I found, in addition to the basic factor of size, first a leptomorphic group factor subdivided into narrower factors for limb length and trunk length respectively, and then a pachymorphic group factor subdivided into narrower factors for trunk girth and limb girth respectively. If we assume that the factor for trunk

¹ Kretschmer stresses *two* features as being combined in what he calls the pyknic type, namely, a tendency to (i) breadth and (ii) rotundity (8, p. 32 : in the English translation the last sentence on this page is a little misleading : " the pyknic stands below the athletic on an average, while he stands above him in chest measurement." This should read : " the pyknic type falls below the athletic type in average (i.e., general) measurements, but the former surpasses the latter in chest measurement," as indeed is shown by the tables to which Kretschmer is referring, viz., Tables IV, II and III.) It would thus seem that Kretschmer has amalgamated two traditional types—the ' respiratory ' or thoracic and the ' visceral ' or abdominal—into the single conception of a pyknic type. The bracketing of the two within one broader category is justified, as factor IIB indicates, provided it does not lead us to overlook the further subdivision demonstrated by factors IIIC and IIID. The subdivided factors, as we shall see in a moment, evince different relations to emotional qualities—differences which with Kretschmer's broader grouping are obscured.

girth (which included traits like shoulder girth and chest girth) depends mainly on trunk breadth, and that the factor for limb girth (which included traits like girth of seat, of thigh, of calf, etc.) depends mainly on subcutaneous fat, the factor-pattern is almost exactly the same as that obtained here.¹

And generally speaking, it may, I think, fairly be claimed for the method (provided, of course, the choice of traits renders its use appropriate) that it not only leads more directly to a unique factor-pattern, but also provides a scheme of factors that is reasonably invariant from group to group, even when the traits themselves are partly changed, and moreover possesses far greater power as an instrument of analysis.² But these points I hope to take up in another paper.

V. CORRELATIONS WITH TEMPERAMENTAL FACTORS

Results. For 253 of the cases included in the above series, assessments for the primary emotions³ were also secured. These permit us to calculate factor-measurements for temperamental types according to the scheme outlined in previous articles (12, 21). The correlations between the measurements for the physical factors and the emotional factors are set out in Table V. Significant coefficients are printed in heavy type. The following conclusions may be tentatively drawn.

I. As regards the general factor of all-round emotionality : (i) There is a slight but significant tendency for boys with bodies smaller than the average for their age to be less emotionally active than the bigger or the better developed. (ii) There is a tendency, still more marked, for the fat boy to be less emotional than the average—more placid and more lethargic. The thin and slender boy tends (if anything) to be slightly more emotional than the average, though here the correlation is barely significant.⁴

II. As regards the bipolar factor distinguishing extravertive from introvertive tendencies : (i) The big, well-developed lad inclines to be more extraverted than the smaller boy. (ii) The leptomorphic type inclines to be more introverted, and the pyknic type (if anything) to be more extraverted, than the average : the tendency is most marked in those who owe their leptomorphic characteristics chiefly to length of limbs and their pyknic characteristics chiefly to breadth of trunk. (iii) The correlation between fatness and extraversion, though positive, is in this investigation non-significant.⁵

III. As regards the euthymic-dysthymic factor : (i) The leptomorphic child tends also to be less happy than the others : this tendency seems more particularly associated with the possession of a trunk that is long, narrow, and relatively deficient in fat. (ii) The broad-built lad and still more the plump youngster tend to be jolly and contented.

¹ Cf. 19, p. 185, Table 4 : cf. also 20, Table 3, and *Man.*, LXXII (1944), p. 85.

² Thus, using ordinary bipolar analysis, we should need at least 2^{m-1} traits to verify m *fundamenta divisionis* ; using subdivided bipolars, we can verify as many as 2^{m-1} *fundamenta* with 2^{m-1} traits.

³ Our earliest attempts at collecting assessments for emotional characteristics were based on the scheme drawn up by McDougall for the Brit. Ass. Anthropometrical Committee (4, pp. 373-4). This, however, was considerably modified along the lines described in later papers (12, pp. 164f., 17, pp. 111f.). The assessments for the present group were determined primarily from ratings and replies furnished by teachers : the standards of different teachers were checked by a personal reassessment of representative children, based on an interview and a situations-test. The reliability and validity of the procedures are discussed in 17. Cf. also 17, pp. 369f.

⁴ In a smaller and more selected group described in a previous study, a larger and fully significant correlation was found between slenderness and general emotionality. But, as was pointed out at the time, "much of the correlation seems traceable to the fact that the group contained several who showed mild hyperthyroid symptoms" (12, p. 186).

⁵ In my earlier enquiry a significant correlation was found between 'plumpness' and extraversion in the small selected group of children and a still larger correlation among the adults (*loc. cit.*, p. 186).

TABLE V. CORRELATIONS OF PHYSICAL WITH EMOTIONAL FACTORS

Physical Factors		Emotional Factors		
		I. General Emotionality	II. Extraversion vs. Introversion	III. Euthymia vs. Dysthymia
I.	Basic Factor (Body Size)	0.154	0.163	0.122
IIA.	Leptomorphic ..	0.112	— 0.134	— 0.155
IIB.	Pyknomorphic ..	—0.096	0.123	0.081
IIA1.	Length of Limbs ..	0.123	— 0.189	—0.053
IIA2.	Length of Trunk ..	0.032	0.113	— 0.167
IIB1.	Breadth of Body ..	0.078	0.145	0.139
IIB2.	Thickness of Body (including obesity)	— 0.230	0.101	0.148

Significant coefficients are printed in heavy type.

Interpretations. At first sight these inferences may seem in the main to sanction popular impressions about temperamental correlates of the so-called 'asthenic' and 'pyknic types,' and their original bodily constitution (8, p. 18). But three important qualifications should be emphasized. First, the figures are far lower than recent statements of the doctrine would imply. Here, as in previous researches, it seems evident that "the measurable correlations, though frequently positive, are too slight to be of value for practical diagnosis" (9, p. 23). Secondly, the usual description of the 'constitution types' needs modification. The chief characteristic of the typically pyknic child is not so much his extravertive or cyclothymic disposition; what he really exemplifies is rather the marked correlation of emotional lethargy with the excessive adiposity that is said to be the distinctive feature of a 'pyknic body-build.' Thirdly, it must not be inferred that the correlations necessarily point to some innate or constitutional linkage of physical with emotional characteristics. Post-natal influences of various kinds would often suffice to explain the facts, and in some cases can plainly be demonstrated.

For example, one or two of the correlations seem undoubtedly to have been enhanced by the inclusion of cases exhibiting minor endocrine peculiarities, especially hypo- or hyper-activity of the thyroid and possibly of the pituitary; and, although in a few instances the endocrine peculiarities may themselves be innately determined, in others they appear to be the result of some mild and intercurrent pathological condition. On the whole, however, I am inclined to think that medical writers have attached too much importance to the 'double influence of the ductless glands' (a type of causation with which they are naturally familiar) and too little to social or psychological agencies. In a number of my most typical cases the history often indicated that the child's physical characteristics were the secondary effect of his emotional characteristics, not vice versa. This was most noticeable with the obese types: in several of these the excessive fat appeared after the child (for reasons easily ascertained) had dropped his former love of exercise and had taken to sedentary habits and to over-eating.¹ Among young males of the age investigated the mere possession

¹ Some of my cases (two or three of them referred to me with a provisional diagnosis of "pyknic, query pituitary disorder") proved to be sheltered and pampered little creatures, who had been allowed no adequate physical enjoyments and consequently found their chief satisfaction in sweets and food. Several of these were readily cured by a change in parental management or, in a few instances, by the change of home and leisure activities ensuing on war-time evacuation. However, a discussion of causes falls outside my present purview. In another paper I hope to present factorial results for a series of age-groups, and to compare growth-tendencies and other causal influences in greater detail.

of a body that is larger and stronger than that of one's companions naturally favours the development of an extravertive and even an aggressive attitude. On the other hand, quite apart from any innate temperamental tendency, it is only natural to find the ill-nourished, narrow-chested youngster less active, less cheerful, more timid and more sensitive than his fellows.

VI. SUMMARY AND CONCLUSIONS

1. It is contended that the absence of systematic structure, which many recent psychologists have reported or assumed in the course of their factorial work, may result from inadequate methods of analysis. With many types of correlation table, instead of extracting group or bipolar factors of the ordinary pattern, it would seem more natural to seek a factor-pattern in which the broader bipolar (or group) factors will be progressively subdivided into two or more independent bipolar (or group) factors of narrower range. Calculations of this kind, it is claimed, would furnish a more plausible interpretation of the correlations often obtained in the course of psychological and anthropometric work, and be more in keeping with the hierarchical schemes of classification and organization, with which mental and other biological characteristics appear to conform.

2. The chief criterion for this type of factor-pattern is the non-significance of the residual cross-correlations which a second general bipolar factor would imply. Instances are quoted from psychological data in which this criterion seems manifestly fulfilled ; and, by way of illustration, a correlation table for a set of physical measurements are analysed in detail by both the old method and the new.

3. 457 boys, aged 11 to 13, have been measured for various bodily characteristics. The correlations obtained have been factorized first by the ordinary method of simple summation. The data have then been factorized afresh by the method of subdivided factors. The results indicate that the so-called leptomorphic or longitudinal factor includes two sub-factors—for limb length and trunk length respectively—and that the so-called pyknic factor similarly includes two sub-factors—for transverse and sagittal measurements respectively (the sagittal measurements being associated with a tendency towards obesity).

4. The calculated measurements for the physical factors as thus determined have been correlated with calculated measurements for the chief emotional or temperamental factors. It is found that several of the correlations are significant, but that all are small.

5. The coefficients obtained suggest several modifications in current theories about the nature of physical and emotional types and their interrelations. The pyknic types appear to include two very different temperamental subtypes : the obese subtype is found to be correlated with lack of general emotionality, the eury-morphic with extraversion. It is pointed out, however, that the correlations do not necessarily imply that body-build and emotional disposition are co-ordinate results of an innate underlying constitutional cause. In some cases (e.g., those showing malnutrition) the correlated emotional disposition would seem to be an indirect effect of the body-build ; in others (e.g., those showing obesity) the body-build may occasionally result from an acquired emotional disposition.

VII. APPENDIX: A SHORTENED METHOD OF FACTOR ANALYSIS

A special advantage of the principle of 'simple summation' is that it permits, and indeed suggests, several useful short-cuts for extracting factors and computing their saturations. Most of them depend on a preliminary 'condensation' of the matrix of observed correlations into a matrix of lower order by first partitioning the initial matrix into a set of suitable submatrices, and then summing the elements in each submatrix.¹ For general purposes the most useful working-procedure is that exemplified in Table V.

After arranging the table of observed correlations according to the more conspicuous clusters, or (it may be) after actually calculating an approximate factorial matrix by the fuller method, the broad lines of division for the first two or three bipolar factors will, as a rule, be fairly clear. The cross-classification furnished by the sign-pattern of these factors will then serve to divide the traits into four distinct groups. Accordingly, we proceed to partition the table of observed correlations into four oblong submatrices, and calculate the sum of the coefficients in each of the short columns that make up the four submatrices: (cf. Table I, p. 49). Instead of 12 rows of 12 coefficients, we thus have four rows of 12 sums. We can now factorize these four rows by the same procedure that we should adopt for factorizing the original table.

In Table VI the four rows of sums are written out afresh and marked O (i.e., observed figures). We first find the total for each column by adding the four appropriate sums. These totals are then entered below in the first line of the lower half of the table, and marked ΣO .

These, of course, are the column-totals of the original correlation matrix; and from these the saturations for the first factor can be computed in the usual way. Each column-total is divided by the square root of their grand total; and the saturations are entered in the row below, labelled Sats. 1. We now partition this row of saturations into the same four groups to correspond with the partitioning of the original correlation table. We then add the saturations for each group, as indicated by the horizontal braces. To calculate the rectangular product-table we multiply the several saturations by these sums (instead of by each saturation taken individually). The four rows of products (H1) are then subtracted from the corresponding rows of observed sums (O). The four rows of remainders (R1) represent the summed residuals.

If the arithmetic is correct, then for each trait the sum of the residuals from the first two rows should be numerically identical with the sum of those from the second two rows. But since the signs of the latter are the opposite of those of the former, we must weight the first two rows of residuals with +1 and the second two rows with -1, before calculating the column-totals for these residuals: (otherwise the column-totals would obviously be zero).

We now use these column-totals ($\Sigma R1$) to obtain the saturations for the second factor (Sats. 2). The procedure is the same as in the full method: weight the column-totals appropriately with +1 and -1 before finding the grand-total (otherwise it would be zero); and divide each column-total with the square root of this grand total.

We sum these saturations in groups as before; and calculate a product-table for the second factor (H2). As before, in summing the residuals the scheme of weighting by ± 1 follows the expected sign-pattern of the next set of saturations. Continuing in this way, we can always extract the saturations for at least three factors. In the present case it is possible to calculate the saturations for four² factors, and that on a single sheet.

At each stage, it will be observed, we have convenient checks on the arithmetic. The algebraic total of each row of residuals should always equal zero: the numerical total of each hierarchical row should always be the same as the numerical total of the corresponding row of residuals from which it is to be subtracted, and both can be checked by multiplying the group-totals of the saturations by their grand total. The practised computer will be able to shorten the procedure still further: the rows of proportional products (H1, etc.) need not be written down explicitly: they can be calculated and subtracted in one process on the machine. The numerical values in certain rows of later residuals are identical: and these too need not be calculated, unless an error is suspected.

¹ Cf. 'Factor Analysis by Submatrices' (13), pp. 339-75 and refs. In what I call the 'sum method' and the 'sum-and-difference method' the original square $n \times n$ matrix is condensed vertically to a rectangular $s \times n$ matrix (i.e., a matrix containing only s rows of n elements, as illustrated in Table I above); in the 'group factor method' it is then condensed horizontally to a square $s \times s$ matrix.

² In the case of the fourth factor, when the divisions between positive and negative signs of the saturations do not correspond to the divisions between the four sections introduced by the second and third factors, special care must be exercised in changing the signs of the column-totals before calculating the grand total of the residuals. If the computer recollects that sign-change is not a mere sign-reversal, but based upon systematic weighting of the horizontal row to correspond with that of the vertical columns, he is not likely to go wrong: an error, if made, can be readily checked in several ways, e.g., the residuals from the fourth factor will not be zero.

TABLE VI. THE SUM METHOD OF FACTORIZATION

Trait	1	2	3	4	5	7	8	9	10	11	12	Check Sum
0	2874	2720	2991	3305	1819	2284	1936	1777	-829	-575	-492	23441
H1	2303	1992	2375	2322	1803	2463	2248	2407	1402	1320	-924	23441 (7817 × 2998)
R1	-571	-728	-616	-783	-016	-179	-312	-650	-573	-765	-432	0000
H2	-343	-448	-378	-740	-461	-293	-258	-657	-578	-691	-038	5742 (4156 × 1382)
R2	-228	-280	-238	-243	-504	-144	-054	-027	-005	-054	-038	0000
H3	-155	-176	-173	-106	-292	-257	-119	-090	-081	-205	-180	2152 (2756 × -781)
R3	-073	-104	-065	-137	-212	-143	-173	-063	-086	-151	-142	0000
0	-887	-763	-927	1081	1508	-748	-845	-629	-279	-247	-218	9608
H1	-943	-817	-974	-091	-1772	1009	-922	-987	-575	-204	-378	9608 (7817 × 1229)
R1	-056	-054	-047	-130	-736	-261	-177	-358	-236	-264	-160	0000
H2	-172	-226	-191	-373	-252	-147	-131	-331	-231	-248	-198	2892 (4156 × -696)
R2	-228	-280	-238	-243	-504	-114	-054	-027	-085	-054	-038	0000
H3	-155	-176	-173	-106	-292	-257	-119	-090	-081	-205	-180	2152 (2756 × -781)
R3	-073	-104	-065	-137	-212	-143	-173	-063	-086	-151	-142	0000
0	1651	1323	1709	1314	1187	2475	2208	2454	1328	1174	-701	18559
H1	1822	1577	1881	1838	1491	1950	1781	1906	1110	1045	-731	18559 (7817 × 2374)
R1	-171	-254	-172	-524	-304	-525	-427	-548	-218	-129	-030	0000
H2	-217	-284	-239	-468	-292	-186	-163	-416	-366	-437	-249	3634 (4156 × -875)
R2	-046	-030	-067	-056	-012	-339	-264	-132	-148	-308	-219	0000
H3	-119	-134	-132	-081	-224	-196	-091	-069	-062	-157	-117	1644 (2756 × -597)
R3	-073	-104	-065	-137	-212	-143	-173	-063	-086	-151	-142	0000
0	-590	-387	-566	-353	-373	-914	-872	-1417	1221	1445	-997	9506
H1	-934	-808	-963	-941	-764	-998	-912	-976	-569	-535	-375	9506 (7817 × 1216)
R1	-344	-421	-397	-588	-391	-084	-040	-441	-652	-910	-622	0000
H2	-298	-391	-330	-644	-403	-255	-224	-573	-504	-602	-343	5002 (4156 × 1203)
R2	-046	-030	-067	-056	-012	-339	-264	-132	-148	-308	-279	0000
H3	-119	-134	-132	-081	-224	-196	-091	-069	-062	-157	-117	1644 (2756 × -597)
R3	-073	-104	-065	-137	-212	-143	-173	-063	-086	-151	-142	0000
Σ 0 Sats. 1	6002 -768	5193 -664	6193 -792	6053 -774	4907 -628	Factor 6421 -821	5861 -750	6177 -803	3657 -468	3441 -440	2408 -308	61114 = 7817 ² 7817
Σ R1 Sats. 2	1031 -248	1349 -325	1138 -274	2225 -535	1388 -334	Factor 881 -212	777 -187	1977 -476	1739 -418	2078 -500	1184 -285	17272 = 4156 ² 4156
Σ R2 Sats. 3	-548 -199	-619 -225	-610 -221	-375 -136	-1031 -374	Factor 907 -329	-422 -153	-317 -115	-287 -104	-724 -263	-635 -230	4156 7596 = 2756 ² 2756
Σ R3 Sats. 4	-294 -120	-416 -169	-259 -105	-548 -222	-849 -345	Factor 572 -232	-693 -281	-252 -103	-344 -140	-605 -245	-568 -231	2756 6068 = 2464 ² 2464
			-781		-616							

Note.—The sums in the check column are calculated after the group-signs have been appropriately reversed. Thus in line 6 the group totals are +610, -610, +466, and -466. The algebraic total is zero. But, on reversing the negative signs, the numerical total is 2152. The total of the preceding line of residuals would be the same if treated in the same way, viz., +989 + (-989) + 087 + (-087) = 2152.

If the reader compares this procedure with the numerous and detailed tables of residuals and proportional subtrahends given in most illustrations of factorial methods,¹ he will see how much is economized in time and labour. It will also be noted that the row-totals calculated by way of preparation for this procedure are precisely the same as those that will be required for the subsequent group factor analysis, whenever that is also contemplated: so that the only piece of extra labour would usually be required in any case. The procedure is particularly useful for obtaining communalities by successive approximation and for checking results in large correlation tables: (consider the saving in checking the saturations for Dr. Cattell's 36×36 table in the preceding issue of this journal, pp. 206-7).

Of the short-cuts based on the summation principle perhaps the speediest is that which we call the 'sum-and-difference' method. With this the bipolar factors are obtained as *differences* between rows of summed residuals—each row being weighted by a figure based on saturations already calculated, which annuls all factors except that which is being calculated at the moment. This is a little too intricate for the beginner or the occasional computer; but will be found particularly useful to experienced and routine computers.²

Note.—In this appendix 'hierarchy' has the derivative meaning, 'matrix of rank one'. In the text it has its original meaning, which may be defined as follows. A 'hierarchical arrangement' (R_h) is a *divergent* order progressively generated by an asymmetrical, *one - many* relation (R), much as a *serial* order is generated by an asymmetrical, *one - one* relation. A 'hierarchy' is the whole class of terms thus arranged (including both the single initial referent and the converse domain of R_h which forms the relata). The n th 'level' is the class of all terms to which the initial referent has the relation R^n . The commonest generating relations are those of 'whole-and-part' (organization) and 'class-inclusion' (taxonomy), yielding 'structural' hierarchies, and of 'conditioning' or 'conditioned by' (e.g., co-ordinating, controlling, using, etc.), yielding 'functional' or 'dynamic' hierarchies.

REFERENCES

1. Galton, F. (1883). *Enquiries into Human Faculty and its Development*. London: Macmillan.
2. Binet, A. (1903). *L'étude expérimentale de l'intelligence*. Paris: Masson.
3. Binet, A. (1905). *The Development of Intelligence*. (Eng. trans. E. S. Kite, 1916. Baltimore: Williams and Wilkins Co.).
4. Committee on Anthropometrical Investigations (1908). *Brit. Ass. Ann. Rep.*, LXXIII, pp. 354-371.
5. Burt, C. (1909). 'Experimental tests of general intelligence.' *Brit. J. Psych.*, III, pp. 94-177.
6. Spearman, C. and Hart, B. (1912). 'General ability: its existence and nature.' *Brit. J. Psych.*, V, pp. 51-84.
7. Burt, C. (1915). 'General and special factors underlying the primary emotions.' *Brit. Ass. Ann. Rep.*, LXXX, pp. 694-6.
8. Kretschmer, E. (1925). *Physique and Character*. London: Kegan Paul, Trench, Trubner.
9. Burt, C. (1927). *The Measurement of Mental Capacities*. Edinburgh: Oliver and Boyd.
10. Spearman, C. (1927). *The Abilities of Man*. London: Macmillan.
11. Thurstone, L. L. (1935). *The Vectors of Mind*. Chicago: University of Chicago Press.
12. Burt, C. (1938). 'The analysis of temperament.' *Brit. J. Med. Psych.*, XVII, pp. 158-188.
13. Burt, C. (1938). 'Factor analysis by submatrices.' *J. Psych.*, VI, pp. 339-375.
14. Burt, C. (1940). *The Factors of the Mind*. London: University of London Press.
15. Burt, C. and John, E. (1942). 'A factorial analysis of the Terman-Binet tests.' *Brit. J. Educ. Psych.*, XII, pp. 117-127, 149-155.
16. Banks, C. (1945). 'Factor analysis applied to current psychological problems.' Doctorate Thesis Univ. London Library.
17. Burt, C. (1945). 'The assessment of personality.' *Brit. J. Educ. Psych.*, XV, pp. 107-121.
18. Cattell, R. (1946). *Description and Measurement of Personality*. New York: World Book Co.
19. Burt, C. (1947). 'Factor analysis and physical types.' *Psychometrika*, XII, pp. 171-188.
20. Banks, C. and Burt, C. (1947). 'A factor analysis of body measurements for British adult males.' *Ann. Eugen.*, XIII, pp. 238-256.
21. Burt, C. (1948). 'The factorial study of temperamental traits.' *Brit. J. Psych.: Stat. Sect.*, I, pp. 178-203.

¹ E.g., in Thurstone's illustration of his procedure (*Vectors of Mind*, Tables 14 to 25, pp. 109-117). And even here his printed tables do not include the tables of proportional products, which all but the most experienced computers would write out in full, in order to check possible errors.

² A fuller account of working instructions is given in *Lab. Notes on Factor Analysis: Sum-and-Difference Method and Sum Method* (ronéo'd).

CORRESPONDENCE

To The Editors, *British Journal of Psychology: Statistical Section.*

Two Criticisms

Sirs, May I venture to criticize two points in your last issue?

1. The first arises out of the conclusions Prof. Cattell draws at the close of his article on 'Personality Factors in Men and Women' (p. 130). His object is to show that the similarity of the factors that he has discovered, first in women and then in men, cannot possibly be due to chance. The evidence turns on the fact that at least *three* out of the *six* 'highest variables' (or traits) are the same in the two factors matched, one from each sex. There are 36 traits in his list. He then argues: "the probability of three given variables out of 36 appearing in any group of six by chance is well below the 1 per cent. level. ${}^{33}C_3$ divided by ${}^{36}C_6$ equals $1/357$." The calculation is correct, but appears quite irrelevant. What he surely meant to estimate was the probability that two groups of six items, both selected from the same list of 36, should by chance alone have three or more items in common. I make this latter probability $2093/46376$, or roughly $1/22$. So far from being "well below the 1 per cent. level," it is only just under the 5 per cent. level. The results *might* therefore quite conceivably be due to chance for some of the personality factors compared.

2. In his generous notice of Dr. Eysenck's book, *The Dimensions of Personality*, your reviewer rightly observes that some of the conclusions, "like the tone of the whole volume, are perhaps a little sanguine." It is, however, disappointing to find that no comment was made on what may be regarded as the most astonishing of all Dr. Eysenck's correlational results.

(a) At the top of p. 83 in his book Dr. Eysenck describes how he standardized his measurements of statures and of transverse chest diameters for 1,000 cases, so that in both cases the distributions had means of 50 and S.D.'s of 10. His measure of Body Size, which was the product of these standardized measurements, had a mean value of 2397. From these figures it can easily be deduced that the product moment correlation coefficient between height and transverse chest diameter for the 1,000 cases must have had the surprising value of -1.03 . Quite apart from the numerical value of the coefficient, the discovery of a *negative* correlation between height and chest breadth conflicts violently with the plain man's notions of anatomical proportions in human beings. From the figure reached it would evidently follow that the shorter a man is, the broader his chest must be, and vice versa.

(b) The factorial analysis from which Dr. Eysenck derives his Index of Body Type is itself based on rather curious anthropometric data. On p. 78 of his book he gives figures for the 200 neurotic patients measured for this purpose. They include:

Arm length to radial styloid: mean 55.7; S.D. 2.78 (in cms.)

Arm length to tip of medius: mean 75.3; S.D. 3.57 (in cms.)

Unfortunately the correlation between the two measurements has here been omitted. It can, however, be found in the original article ('A Factorial Study of some Morphological and Psychological Aspects of Human Constitution,' *J. Ment. Sci.*, 1945, XCI, p. 11). It is stated to be $+0.266$.

For simplicity let us call these two lengths 'arm' and 'arm-plus-hand' respectively, and their difference 'hand length.' Then a little arithmetic will immediately reveal that, in these 200 persons, 'hand length' had a mean value of 19.6 cms. and an S.D. of 3.9 cms. The range for 200 persons would almost certainly extend beyond the limits -2 S.D. and $+2$ S.D., which implies that the hands must have varied at least from 4.6 inches to 10.8 inches! Further the correlation between 'hand' and 'arm' lengths turns out to be -0.47 —another negative correlation, which here implies that long arms tend to grow short hands and vice versa. Nevertheless the Index of Body Type so derived succeeded in giving a correlation of $+0.96$ with that of earlier workers.

My picture of an imaginary visit to the wards of the Mill Hill Emergency Hospital now reminds me of a stroll through the gallery of distorting mirrors at a fun-fair. I see all the tallest men with ludicrously narrow chests, and those with short arms waving hands that are nearly a foot long; and I begin to wonder whether the embarrassment induced by these peculiar anatomical proportions may not have contributed to the neuroticism of the unfortunate patients. It is astonishing that not only your reviewer, but the numerous medical and other colleagues who collaborated in the book, failed to detect these oversights. If such obviously erroneous figures are calculated for everyday features like the measurements of the human body (where the absurdities should strike the most hasty or careless investigator), is not one's confidence bound to be shaken in the remarkably optimistic results obtained with psychological measurements? Yours sincerely, PHILIP D. GREENALL.

Reply. I am indebted to Mr. Greenall for drawing my attention in such a charming and friendly way to two misprints. The second of these occurred in the article quoted by Mr. Greenall: the correlation between the two measurements is ± 0.866 , not ± 0.266 as stated. The other misprint I cannot correct, as the original sheets were lost when the Department moved from Mill Hill to the Maudsley.

I find his references to "remarkably optimistic results" rather puzzling. By and large, our correlations are no higher than those obtained by other investigators; in the sphere of body build, for instance, hardly any of our correlations between temperament and constitution are above the .3 level, which may be compared with Sheldon's results which exceed the .8 level. Similarly, in the field of questionnaires, our results are neither more nor less favourable than typical results from recent American work. The same is true of most of the other tests we used; and it may interest Mr. Greenall to know that since the publication of *Dimensions of Personality* we have repeated much of this work with children, only to find that our results there are even more encouraging. H. J. EYSENCK.

Pearson's Method Modified to Allow for Non-Significant Factors

Sirs, In his suggestive article, explaining how the common sense method of calculating 'factor measurements' by simple averaging leads on to Pearson's method of 'principal axes' (this *Journal*, I, pp. 5, 19) Sir Cyril Burt has expressly restricted himself to the case where the computer inserts in the leading diagonal the total test-variance taken as unity. Yet, in his actual researches, he always inserts, not unity, but what is now usually called the 'communality'; and further in his earliest papers he bases his factor analysis, not on the observed correlations as they stand, but on a "reduced correlation table, corrected for attenuation due to specific factors." He gives a formula (e.g., this *Journal*, I, p. 189): but he does not tell us how it is reached, beyond the rather vague remark that it can "readily be deduced by inserting the supplementary assumptions into Pearson's proof." Could he briefly state what these 'supplementary assumptions' are, and perhaps outline the proof of his 'modified equation'? R. F. CONINGTON.

Reply. The 'modifying assumptions' will vary according to the way we envisage our essential problem: e.g., according as our aim is to distinguish significant from non-significant factors, general (or common) from specific (or unique), relevant from irrelevant, systematic factors from error factors, and so on. Pearson's purpose was simply to express n correlated sets of observed trait-measurements in terms of n uncorrelated sets of hypothetical factor-measurements, every factor being assumed to be common to all the traits, and each being defined in turn as the 'line of closest fit' to the observed measurements or their residuals. This is to seek the *largest* number of common factors; and takes no account of errors or factors specific to each trait. My own view is that, in accordance with 'Occam's principle,' we should seek rather the *smallest* number needed to fit the observed measurements, within the limits set by the standard error—one general factor only, if that proves sufficient.

But this involves assuming, in addition to (say) m significant common factors ($m < n$), n supplementary factors, since these latter must comprise (and in certain cases may be

reduced to) a factor specific to each trait. Adopting the usual notation this means that we shall assume, not $X=FP$ (where P is the $n \times N$ matrix of factor measurements), but

$$X = F_c P_c + F_u P_u \quad (\text{i})$$

where P_c is the $m \times N$ matrix of significant common factor measurements, and P_u is the $n \times N$ matrix of measurements for the supplementary factors, which in the simplest case will each be 'unique' or specific to one trait only. All measurements are assumed to be in unitary standard measure: so that the matrix of correlations, with unity in its diagonal, will be

$$R = XX' = F_c P_c P_c' F_c' + F_u P_u P_u' F_u' = F_c F_c' + F_u F_u' = R_c + R_u \quad (\text{ii})$$

Now we desire to estimate the measurements for each common factor by means of a weighting equation

$$\bar{p}_i = w_i' X \quad (\text{iii})$$

where $\bar{p}_i$ denotes the estimated measurements for the i th common factor. Its variance will be

$$\bar{p}\bar{p}' = w' X X' w = w' R_c w + w' R_u w \quad (\text{iv})$$

Here $w' R_c w$ will evidently denote the contribution of the significant factors to the variance and $w' R_u w$ the contribution of the residual factors. Our aim is to maximize the contribution of the former, or (what amounts to the same thing) to minimize the proportion of the variance assigned to the latter. Accordingly we can either take logarithmic differentials, or, putting $\bar{p}\bar{p}' = \text{constant}$ (say 1), we can adopt the familiar Lagrange device.

Following Pearson, let us introduce an undetermined multiplier, λ . We can then take the expression $w' R_c w + \lambda w' R_u w$; differentiate it with respect to each of the elements in the column-vector w ; and set each of the derivatives equal to zero. We thus obtain

$$w'(R_c - \lambda R_u) = 0 \quad (\text{v})$$

that is,

$$w' R_u^{\frac{1}{2}} (R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}} - \lambda I) R_u^{\frac{1}{2}} = 0$$

Writing $R_a = R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}}$, I call R_c the 'reduced' correlation matrix and R_a the 'augmented' matrix: R_a is, in fact, the reduced correlation matrix 'corrected' for the effect of the unwanted factors, i.e., 'for specificity' if these are assumed to be specific. In this last case the procedure is equivalent to measuring each trait in terms of the s.d. of its own specific factor. The condition for a non-trivial solution is

$$|R_a - \lambda I| = 0 \quad (\text{vi})$$

This equation differs from Pearson's simply in substituting R_a for R . And this in turn means that, as in any least squares procedure where the observed measurements are of 'unequal precision,' we first weight each set of measurements by dividing it by its standard error, so that the 'errors' of each are reduced to the same relative value.¹ (The calculation of 'partial correlations' is based on the same principle.) When $m=1$, i.e., when there is only one significant common factor, the procedure (as a little algebra will quickly show) simplifies to that of weighted summation applied to the reduced correlation matrix; and the same simplification may be adopted for most practical purposes, when we can reasonably assume that the variance of each specific (or residual) factor is approximately the same.

If we write

$$R_a = F_a F_a' = L_a A_a L_a' \quad (\text{vii})$$

we have

$$F_c = R_u^{\frac{1}{2}} L_a A_a' \quad (\text{viii})$$

and

$$F_c' R_u^{-1} F_c = F_a' F_a = A_a \quad (\text{ix})$$

(a diagonal matrix); and, if the estimated factors are required to be in standard measure, we have

$$W' = A_a^{-1} (1 + A_a)^{-1} F_a' R_u^{-\frac{1}{2}} \quad (\text{x})$$

The obvious working procedure is to estimate $\hat{R}_u$, and then factorize R_a by weighted summation, using (viii) to obtain the estimated saturations F_c , and hence improved estimates of R_u . To test the significance of a doubtful factor, we can apply any of the usual tests to the residual correlations 'corrected' for non-significant factors or for specificity: that is, we test, not the ordinary residuals, but the partial correlations. A fuller explanation with examples is available in *ronco'd Laboratory Notes on Factor-Analysis* (1934). CYRIL BURT.

¹ Cf. Johnson, W. W., *Theory of Errors and Method of Least Squares* (1892), pp. 98f. As a consequence, it will be noted, the factors obtained will be the same, regardless of the way the traits are measured, whether in some arbitrary unit of measurement or the standard deviation (e.g., whether based on covariances or correlations).

BOOK REVIEWS

Rank Correlation Methods. By M. G. Kendall. London: Griffin and Co. 1948.
Pp. viii + 160. 18s.

Until quite recently rank correlation was, as Mr. Kendall says, "a rather neglected branch of the theory of statistical relations." In the present volume his primary aim is to "give an account of the new ranking techniques." He adds that he was "encouraged to write it by some of my friends who favour the methods for psychological work." The chief outcome of his discussion is to demonstrate that rank correlation is no longer to be regarded as a somewhat dubious makeshift, but leads to procedures which are of considerable practical value and to problems of great theoretical interest.

He begins by pointing out that ranking has the special merit of remaining "invariant under stretching of the scale." He then deduces two formulæ for rank correlation. Let us suppose that we have n individuals ranked in order for two different traits. We may measure the correlation between the two by one or other of two expressions:

$$(i) \quad \rho = 1 - \frac{3 \sum d^2}{\frac{1}{2}(n-1)n(n+1)},$$

where d is the difference between the two ranks allotted to the same individuals;

$$(ii) \quad \tau = 1 - \frac{2(n+1)s}{\frac{1}{2}(n-1)n(n+1)},$$

where s is "the minimum number of interchanges between neighbours needed to transform one ranking into the other."

The first equation is simply the ordinary Pearson product-moment formula, applied to ranks instead of to ordinary measurements, and appropriately simplified: Mr. Kendall names it after Spearman, who, he adds, "introduced it into psychological work."¹ The second has been ingeniously devised by Mr. Kendall himself (*Biometrika*, XXX, 1938, p. 81). I have ventured to introduce a superfluous term into both numerator and denominator to show how the two expressions resemble and differ from each other. It is not difficult to show that the correlation between the two is approximately $1 - 1/4n$, and that, when n is large, τ will be about two-thirds the value of ρ . Evidently $\sum d^2$ will be easier to compute than s ; and, although there are short cuts, ρ would seem to be the more convenient for practical calculation. On the other hand, τ is the only coefficient of correlation that depends solely on linear processes. Hence a machine might readily be constructed to compute it.

After three introductory chapters on the general theory, Mr. Kendall discusses methods of testing significance. Here, if the practical advantage still lies with the older coefficient, the theoretical advantage undoubtedly lies with the newer. The sampling variance of the newer is of an order twice that of the old; but the old presents considerable difficulties in its sampling distribution.

Mr. Kendall next considers the problem presented when we have a number of rankings, e.g., when m observers rank n objects. As he formulates it, his problem would appear to be that of finding a general factor for a set of rank 'correlations between persons.' He points out that the most obvious procedure would be to average all the possible correlations; but shorter methods can be substituted. And he then shows that, if we rank the items according to the sums of the ranks allotted to the individuals, the ranking so obtained "gives a best estimate in a sense associated with least squares"—the best estimate, it may be added for what the psychologist would call the 'general factor.' Here again " ρ is a more convenient

¹ In reviewing Spearman's article Binet claimed that he and Henri had already used this procedure; and in early psychological discussions the name "Spearman's rank correlation formula" was usually understood to refer to the still shorter procedure given in *Brit. J. Psych.*, II, p. 93. Spearman himself explained that "in the simplified form the standard or r method [Mr. Kendall's ρ] shows itself to be distinguished from our short or R method by the fact that the differences are squared." The simplification of the Pearson equation he attributed to Lipps. However, the popularity of rank correlations in psychology, if not their actual introduction, is undoubtedly attributable to Spearman.

coefficient than τ "; and the expression reached seems identical with that already used by psychologists. (It has been called the "corrected square sum ratio," *Factors of the Mind*, p. 275.)

Mr. Kendall next proves that τ can be generalized to the case of partial rank correlation. The advantage now lies with the new procedure. This, as he proves, yields an expression for the partial coefficient that is formally the same as that for the ordinary partial product-moment coefficient. Finally he deals with the more general problem of 'paired comparisons'—a method which enables us to obtain coefficients of consistency and of agreement for those cases in which the underlying variable is not necessarily linear.

Readers who are acquainted with the chapter on rank correlation in Mr. Kendall's *Advanced Theory of Statistics* (I, pp. 388–437) will be interested to learn that the new account contains much new material and several new tables. The exposition too introduces a novel and helpful device: throughout the volume the author writes in alternate chapters; with each fresh topic the first chapter describes the main results and their applications in simple terms; the next gives the more technical proofs for the equations employed. As Mr. Kendall employs it, the method seems eminently successful—far better than the commoner practices of incorporating the advanced mathematics in the general text or relegating them to footnotes or appendices. The discussions are a model of lucidity; and both the new book and the new correlational procedure may be strongly recommended to the notice of all psychologists interested in developing the statistical side of their work.

C. B.

A Guide to Mental Testing. New edition. By R. B. Cattell. London: University of London Press. 1948. Pp. xvi + 411. 25s.

Prof. Cattell's well known *Guide* has long been recognized as almost the only handbook of tests which covers, not merely intelligence and educational attainments, but all the more important aspects of personality. For the new edition, most of the chapters in the 1936 issue have been enlarged and revised. The previous edition began by accepting Spearman's two-factor theory: our "abilities," it was said, consist of one general factor called 'g,' and a host of specific factors or 's's' (p. 1). The new edition substitutes the three-factor theory, viz., that "abilities may be conceived of as (1) a general ability, entering into almost all performances, (2) certain group factors, covering an area such as verbal, spatial, numerical, musical, etc., performance; and (3) certain abilities which are absolutely specific to one performance" (p. 6). In passing it is a little surprising to see this hypothesis attributed to Spearman and Thurstone, since Spearman's two-factor was expressly intended to exclude group factors, and Thurstone as late as 1938 declared that "so far, we have not found a general factor."¹

The chapters on tests of temperament and character are now based mainly on Prof. Cattell's own suggestive work on the measurement of personality; and chapters V and VI open with a very clear statement of the scheme of 'primary personality factors' that he has put forward. It is disappointing therefore to find that, in the 'Notes on Mathematical Formulæ,' the brief section dealing with 'Factor Analysis' still treats only of the 'tetrad difference' method and the "division of abilities into a general factor and factors specific to each test," that is to say, with the discarded two-factor theory only. A suggestive chapter on 'The Measurement of Intra-Familial Attitudes' and two new appendices giving a 'Questionnaire for Personality Factors' and a 'General Information Test' have been added; and the collection of new and old tests now furnish a most valuable source of material not only for the practical psychologist in the child guidance clinic but also for those engaged on research.

J. W.

¹ Thurstone, L. L., *Primary Mental Abilities*, 1938, p. vii. Since Cattell's statement of the 'three factor theory' (including the four group factors explicitly named) seems taken almost verbatim from the Board of Education 1924 Report on *Psychological Capacities* (p. 19), the theory as he cites it can hardly owe anything to Thurstone; certainly it was strongly criticized by Spearman two years later (*Abilities of Man*, p. 241).

ITEM DIFFICULTY AS THE MEASURING DEVICE IN OBJECTIVE MENTAL TESTS

By E. A. PEEL

University of Durham

I. *Introduction.* II. *The Nature of Measurement.* III. *Measurement by Objective Mental Tests.* IV. *Item Difficulty as a Measure in the Sense of Physical Measurement.* V. *Conclusion and Application.*

I. INTRODUCTION

Before the war the British Association appointed a committee (1) to report on the measurement of sensory discrimination, as carried out by the familiar psychophysical procedures. I want now to consider another kind of psychological measurement, that attempted by the use of mental tests. This might be called true psychological measurement, since the stimulus is no longer measured in the units of physics.

In examining any kind of measurement we have to consider two problems, the construction of the measuring device and the accuracy of the measurement achieved. The first leads to the fundamental concepts of measurement; the second seems best approached by considering the errors of estimation, as recently reconsidered by Ferguson (2). In addition when we consider psychological measurement, we have also to examine the problem of validity,¹ that is, the accuracy with which a test predicts the psychological criterion it purposes to assess (3, p. 198).

In this paper, I shall confine myself to the problem of constructing a psychometric device, the objective mental test, and examining its relation to the measuring

¹ Strictly speaking the problem of validity is not confined solely to psychological measures. Mass and length are in large part defined by the operation of measurement, and carry little meaning outside this operation. When we consider mass as the result of an operation on matter, we see that it may be no more real than intelligence. Modern physics has made us aware of the limitations of these measures as predictors when they are applied under conditions outside the usual experience of the senses. Thus we now know that mass is not absolutely constant or unidimensional but must be considered in conjunction with energy and that length is associated with time. Psychologists are made more aware of the need for predictive value in a measure since their measures rarely, if ever, approach unidimensional properties; and so, when applied to individuals, lack the constancy which is associated with physical measures under normal conditions.

Psychologists are often asked to explain what they measure, e.g., intelligence, whereas the measures of the physicist pass unchallenged. There are at least three reasons for this. Intelligence as defined by the simplest of tests is not so obviously unidimensional as mass or length. Also the measure of intelligence is applied to beings who can introspect and express a point of view. Lastly, it seems apparent that the measures which have been developed in physics possess what might be called *relevance*. Relevance is not equivalent to validity; a test may be valid in relation to a given criterion but the same test may not be relevant to a criterion closely related to human practice and experience. If a measure was defined solely by the operation of measurement, we could theoretically define a new measure having little relation to normal experience, e.g., entropy. This is a serious problem in psychology, for most lay criticism of intelligence testing concerns its relevance to what people experience as intelligent behaviour. Otherwise, a psychologist might define intelligence by a particular test, provided of course that he could fulfil the condition that the measure was stable.

However, in actual practice, the psychologist is limited in his choice of a measure of intelligence by a choice of the situations which he and society consider to reveal intelligent behaviour, and hence the measure which results is more or less relevant.

devices of physics, with a view to determining how far and under what conditions the objective test conforms to the requirements of such measurements.

II. THE NATURE OF MEASUREMENT

Measurement concerns the comparison of classes of individuals. Its first requirement is that the individuals to be measured should be capable of classification in respect of an abstracted quality. Secondly, the members of the class should be capable of an order which has some meaning (4, p. 115 and 5, p. 86). The individuals constituting it must conform to certain logical conditions: the connexive, transitive, and asymmetric postulates (see Campbell, 6, pp. 267–270, and Burt, 7, p. 118, for a comprehensive treatment). The third requirement is that the order should be correlated with the members of a number class: the numbers used as measures should be in the same order as the quality measured.

If these three requirements were sufficient for the measurements of physics and mensuration, we might not need to pursue the enquiry further, for much of psychological and educational measurement (assessments or rankings) requires no more than the concepts of class order and correlation. Indeed many psychologists—see Burt (7 pp. 115–137) and Brown and Thomson (8, pp. 11–12)—would be satisfied with these conditions, provided that something more were known of the *difference* between the ranks or points.

However, more serious divergences between physical and psychological measurement occur where the physicist requires that measurement shall be in terms of units which are *constant* and *additive*. We are so accustomed to the additive quality of length and mass and the constancy of their units that we often fail to appreciate that this requirement is arrived at experimentally by a physical *operation*¹ and that it is not a logical process. The construction of a measuring instrument is an experiment from which it is highly probable that constant additive units will be obtained. Campbell (6, p. 267 and 1, p. 340) holds that this requirement can be fulfilled only where the physical *act* of adding (placing end to end) can be realized.² The psychologist may counter this by stating, as Burt (7, p. 117) and Guilford (9, p. 3) do, that any physical measurement is in the last analysis a psychological process of observation. None the less, in my opinion, the psychologist cannot escape this issue of constancy and addition of units. Even if he objects to the physicist's notions of operation he must find some parallel device by which his units may approach constant and additive properties. The construction of a psychological measuring device is also an *experimental* operation; and for this reason the logical postulates of measurement will not by themselves suffice to give an adequate measure. However, obviously the operation of placing end to end cannot be realized with *mental* phenomena.³

¹ Also that the act of measurement is an experimental *operation*, e.g., the application of a test changes the person, producing the learning effect during the process of testing. Similar problems arise in physics in measurement of inter-electronic distances; electrons are pushed apart by the act of finding the distance between them.

² In the light of this condition it is not clear how Campbell would deal with the now known fact that mass and length are not unidimensional quantities: *vide* his distinction between qualities and quantities (6, p. 280); a quantity is changed by more of a substance, whilst a quality is not.

³ R. J. Bartlett (10, p. 423) in his 1939 presidential address to Section J of the British Association stated that we might need to redefine measurement in psychology since psychological measurement cannot conform to this requirement. I shall hope to show that the differences between psychological and physical measurement are not so marked as Bartlett implied.

Guilford observes (9, p. 2) that "the constancy, and therefore the dependability of the unit, is the prime requirement of all measurement." I shall endeavour to show how this constancy of unit might be realized in at least one type of mental test through the notion of scaling the test items according to difficulty and assuming that difficulty is unidimensional.

The above requirements of measurement can be easily discerned in the simple physical experiment of finding the lengths of a set of books. By it we compare two classes of objects, books and straight edges (rulers), in respect of a common abstracted property 'length.' 'Length' is found experimentally to be capable of addition, that is, it has a zero point. Such measurement is said to be direct. We may consider other kinds of measurements, e.g., biophysical, psychophysical, and psychological gradings, and measurement by objective mental tests: the fundamental differences, if any, between these methods would seem to depend on the nature of the classes of individuals correlated to produce the measurement and the degree to which they possess the additive property.

Thus, in biophysical and psychophysical measurement we compare a class of beings or a class of objects with additive qualities to which numbers are assigned in order of magnitude. In biophysical measurement the abstraction from the class of beings is anatomical; in psychophysical it is concerned with mental responses (1). Examples are the measurement of sensation limens and judgments of lengths. Only *one* of the classes, the objects, possesses an additive property; and the response of the person is *measured* in terms of this property. Such measurement is indirect.

In subjective-objective psychological grading we compare a class of beings with a class of subjective opinions or impressions to each of which a number or letter is assigned to give a so-called scale. The increments represented by the numbers are not necessarily equal. If the assessor is not merely concerned with ranking the persons within their class, but also in comparing them with others he has experienced, he is really comparing one class of beings with another in respect of the impressions they produce on the assessor. These impressions are partly subjective and partly objective, depending upon the nature of the abstracted 'impression' or 'production' which is used as the basis of comparison. Brown and Thomson (8, pp. 11-12) and Burt (7, pp. 130-131) outline briefly how such gradings may approximate to the conditions of physical measurement.

III. MEASUREMENT BY OBJECTIVE MENTAL TESTS

In considering measurement by mental tests I shall take a test whose items have been already evaluated for difficulty. It will be shown that, at least in this type of test, we may come very close to realizing constancy of units and the additive law of measurement. Thorndike (12, p. 469) says that measured intelligence may show itself in three ways: in altitude, i.e., level of difficulty attained, in breadth, i.e., proportion correct of items of given difficulty, and in speed, the number of correct responses in a given time. If item difficulty is evaluated as subsequently suggested, it may be regarded as a *measure* of altitude.¹ If its estimation is made by experts, then measurement by objective tests whose items have been so evaluated differs little from measurement by rating scales, save that the several responses are scored objectively. However, in order that its units may approximate to constancy, item-difficulty can be assessed more satisfactorily by counting the correct and incorrect responses of a population to whom the test is applied prior to its use for measurement.

In construction such a test is composed of a number of 'tasks.'² These may be simple arithmetical operations, verbal items, or require the detection of a fault

¹ It may be possible to treat tests of breadth and speed in measured intelligence in a similar way but this problem is not considered in the present paper.

² Many psychologists would select a large number of such items in as random a manner as possible and proceed no further, relying on the statistical probability that, when a test contains a large number of such items, the increments of difficulty become very small and approach equality (see Culler, 13). This is, however, not so satisfactory as the complete process of preliminary evaluation of difficulty.

in a pattern, etc. They can be scored objectively, right or wrong; and yield the unit of measurement. Though differing qualitatively, casual inspection shows that some are easier to perform than others. In this property of item difficulty we may perceive a unidimensional variable.

The preliminary test is given to a large sample to find the degree of difficulty of each item. This phase corresponds to the 'operation' of setting up a physical measure: i.e., it involves the comparison of the class of test situations with a class of persons, from whose mental qualities we abstract their response to the given test situations. To each item we assign a number corresponding to the number who fail to respond to it correctly. This gives its difficulty. These numbers do not necessarily increase linearly. But we may then select the items so that the array does increase linearly in difficulty. The test is then assembled so as to begin with the easiest item and end with the most difficult item. In this way the numbers assigned to the different items for their difficulty, have a similar meaning to those assigned to the inches on a foot rule, where the *position* of any inch on the rule gives the number assigned to that inch.

The test is now ready to apply as a measure in the sense of Thorndike's 'altitude' or as a 'power' measure (see Vernon 14, p. 24). It is given to a person and the item at which he ceases to answer correctly is taken as a measure of his ability. Actually he will often make mistakes before reaching his limit; in such cases we add the number of correct responses. These deviations tend in a well-constructed test not to depart widely from the simple type of answer pattern. Nevertheless, their existence constitutes a divergence from the kind of measure which is used in physics. They are considered by Ferguson (2, Ch. VI).

The final application of the test consists of a comparison of a class of persons with a class of test situations, in which each situation has a number attached signifying its difficulty. Thus we are really *comparing one class of persons with another in respect of a common abstracted property*—'response to the test situations.' In this respect the measure is *direct*, much as a physical measurement like length is direct, save that, instead of objects, we are comparing classes of persons. *We are thus maintaining the general conditions of measurement and at the same time using a genuine psychological property for the measuring device.*

IV. ITEM DIFFICULTY AS A MEASURE IN THE SENSE OF PHYSICAL MEASUREMENT

We now enquire under what conditions the units may be constant and possess a zero point. We can demonstrate the process of finding item difficulty by using physical objects and measuring scales instead of persons and test situations. We can then discover the conditions under which the estimated lengths of scales agree with the 'real' lengths which we already know. Let us compare a class of cardboard strips with a class of objects in respect of length.

Suppose we have 5 strips, 12, 10, 8, 6, and 4 in. long, and 60 objects—5 at 12 in., 5 at 11 in., and so on to 5 at 1 in. We lay each strip on all the objects and assign a number to each corresponding to the *number* of objects which it equals and exceeds.

Real length of strip	12	10	8	6	4	inches
Number of objects equalled and exceeded	60	50	40	30	20	units

The estimated lengths of 60, 50, 40, 30, and 20 are evidently measured in units which are constant

and possess the same zero point as the real lengths.¹ They are in every way equivalent to the real lengths, except for the new unit.

In a second experiment, let us take 60 objects distributed as follows :

Length ..	12	11	10	9	8	7	6	5	4	3	2	1	inches
Frequency ..	1	2	5	5	6	10	13	7	4	4	1	2	total 60

Proceeding as before we obtain the following estimates :

Real length ..	..	..	..	..	..	..	12	10	8	6	4	inches
Estimated length ..	..	..	..	..	..	..	60	57	47	31	11	units

The estimated lengths are now not linear ; and we see that the use of a non-linear distribution of objects would lead to false estimates.

For the last example we may suppose the objects are distributed as follows :

Length ..	12	11	10	9	8	7	6	5	4	3	2	1	inches
Frequency ..	12	12	0	0	12	0	12	0	12	0	0	0	total 60

This gives the following estimates :

Real length ..	..	..	..	..	..	..	12	10	8	6	4	inches
Estimated length ..	..	..	..	..	..	..	60	36	36	24	12	units

Over the range of real lengths 8, 6, and 4 in., the estimated lengths are linear, but have not the same zero point, as the ratio test shows. The displacement of the zero point is caused by the zero frequency in the 2 in. and 10 in. objects.

Estimation of a strip length by the number of objects which it equals and exceeds is analogous to finding the difficulty of an item by counting the number who fail to answer it correctly.² We may therefore embody the above results in a general derivation of the conditions necessary for the units of item-difficulty to be constant and possess a zero point. Let us apply the items to a population P of persons whose abilities are ranged over k consecutive integral classes, k representing greatest ability and 1 least. If there are p_i persons in the i th class, then $P = \sum_{i=1}^k p_i$; and we have—

Person ability : $k, k-1, k-2, \dots \dots \dots 2, 1$ units

Frequency : $p_k, p_{k-1}, p_{k-2}, \dots \dots \dots p_2, p_1$. Total P .

Suppose we have items possessing the same increments of real difficulty as the above abilities and that the lowest ability corresponds with the easiest item. The estimated difficulty of the item of difficulty m (failed by all persons up to ability m) is given by $\sum_{i=1}^m p_i$. Similarly, the estimated difficulty of an

item of difficulty n is given by $\sum_{i=1}^n p_i$. The ratio of the estimated difficulties m and n is given by

$$r_{est.} = \frac{\sum_{i=1}^m p_i}{\sum_{i=1}^n p_i}.$$

The condition for constancy of units and a common zero point is given by $r_{est.} = \frac{m}{n} m/n$.

In general this is only true when $p_i = p_j = p$ say, i.e., when $r_{est.} = \frac{mp}{np} = \frac{m}{n} mp/np = m/n$.

The distribution of P must therefore be rectangular in order that the estimates of difficulty shall correspond with the real difficulty.

We have assumed that the class intervals of person ability and item difficulty are equal. Now consider the consequences of using a scale of abilities whose classes do not coincide with those of item difficulties. This we can do by making zero the frequencies p_i which correspond to those ability

¹ The test of constancy and a common zero point is given by the ratio : $\frac{a_{est.}}{b_{est.}} = \frac{a_{real}}{b_{real}}$, e.g., $\frac{50}{20} = \frac{10}{4}$.

² Subject to similar observations as were made in the previous section on divergences in answer patterns. The physical analogy of these divergences could be realized by the repeated use of objects whose dimensions were successively increased and decreased by distortion between each occasion when they were used to estimate cardboard lengths. Thus a given object might exceed a given strip on one occasion and fail to exceed it on another and so on.

classes we wish to delete. Take first the case in which the ranges of ability and difficulty are the same but in which the ability intervals are larger than the difficulty intervals. This we can effect as follows, Suppose k is even and $p_2 = p_4 = p_6 = \dots = p_{2t} = p$ and $p_1 = p_3 = \dots = p_{2t-1} = \dots = 0$. Then the intervals of person ability are twice as large as those of item difficulty; but the extension

to any other factor than two is obvious. We then have $r_{est.} = \sum_1^m p_i / \sum_1^n p_i = \frac{m}{2} p / \frac{n}{2} p = m/n$; and

the condition for *constancy of unit* and a *common zero point* is fulfilled. It is therefore unnecessary for the units of person ability and item difficulty to be the same. The use of a larger ability unit, however, would cause the item difficulty scale to be less discriminating.

Now consider the case in which the range of person ability does not extend *down* as far as the lowest level of ability. This we can realize by making $p_t = p_{t-1} = p_{t-2} = \dots = p_2 = p_1 = 0$ and the remaining $p_t, t > j$, equal to p . Then, by the above equation, we shall have

$$r_{est.} = \frac{(m-j)p}{(n-j)p}$$

Evidently this expression is not equal to m/n and so the estimated units, although constant in virtue of $p_k = p, k > j$, have not a common zero point with the real values of difficulty and so are not additive.

We may summarize these findings as follows. In order to obtain constant and additive units of estimated item difficulty—

- (a) the distribution of persons used for evaluating item difficulty must be rectangular;
- (b) the lower limit of the range of person ability must extend to zero ability; the upper limit is determined by the most difficult item used;
- (c) the units of person ability need not be as fine as the units of item difficulty; but by making the gradations of ability finer, a more sensitive scale of item difficulty is obtained.

V. CONCLUSION AND APPLICATION

Measurement entails logical postulates of classification common to *all* kinds of measuring operations. These operations are not exact for any science; but success is highly probable in the physical sciences. Physical measurement entails physical acts, such as laying lengths end to end. Mental measurement involves units which have a *psychological meaning*, such as the difficulty of tasks, speed of work, range of tasks, mode of response when ability and difficulty are held constant and so on. Given this, the only other condition we have to fulfil is that the units shall be as constant and as additive as we can *make* them.

Psychological measurement, with objective mental tests, whose items have been previously evaluated for difficulty, consists of a comparison of two classes of persons in respect of their response to the test situations. The comparison constitutes a direct measurement.

The abstracted quality 'response to the test situations' implies the use of 'test situations' as a metric device; and this device is analogous to physical measurement in so far as 'item difficulty' conforms to the same conditions. The estimated item difficulties tend to be linear and additive when derived from a rectangular distribution of persons whose range of ability has a lower limit of zero.

Practical difficulties arise with the selection of the rectangular population. For intelligence tests, its range should extend from the lowest known forms of idiocy. This condition may be substantially realized if we adopt the assumptions made by Thorndike (12, p. 54) and McCall (15, p. 272 f.). If we assume with McCall that mental ability is distributed normally and furthermore that $\pm 5\sigma$ covers almost the entire observable range, we may avoid any difficulties entailed in the selection of a rectangular population, and, since the normal curve is continuous, work with a population which is infinitesimally graded.

We should now apply the preliminary form of the test to a large random sample of persons, allowing each person sufficient time to attempt all items. We could select any group uniform in age and relate other ages to the absolute measure by applying Thurstone's method (16) for scaling tests. Then for each item we should find the proportion p of the persons who fail to answer the item correctly and then from the normal tables we should find the σ value x , i.e., the deviation in standard units, corresponding to each value of p . If we assume the lower limit of ability is -5σ the *absolute item difficulty* for each item is given by $(5 - x)\sigma$ units.

The principles of the method outlined in the previous paragraph could be applied to convert any population distribution to a rectangular form, thus making it possible to convert apparent item difficulties into true values. This of course could only be achieved if the distribution had a *known* functional relation with a linear series.¹

The items would then be reassembled in order of their absolute item difficulty and such items removed as were necessary to form a linear scale. The test would now be in a form for application to obtain a persons absolute ability as measured by counting up his correct responses.

REFERENCES

1. *Brit. Ass. Ann. Rep.* (1939-40). *Final report of the committee on the possibility of quantitative estimates of sensory events*, pp. 331-349.
2. Ferguson, G. H. (1947). *The reliability of mental tests*.
3. Chambers, E. G. (1943). 'Statistics in psychology and the limitations of the test method.' *Brit. J. Psych.*, XXXIII, pp. 189-199.
4. Russell, B. (1927). *Analysis of matter*.
5. Jeffreys, H. (1937). *Scientific Inference*.
6. Campbell, N. R. (1920). *Physics—the elements*. Chap. X.
7. Burt, C. (1940). *The factors of the mind*.
8. Brown, W., and Thomson, G. H. (1940). *Essentials of mental measurement*.
9. Guilford, J. P. (1936). *Psychometric methods*.
10. Bartlett, R. J. (1939-40). 'Measurement in Psychology.' *Brit. Ass. Ann. Rep.*, pp. 422-441.
11. Finney, D. J. (1947). *Probit analysis*.
12. Thorndike, E. L. (1927). *The measurement of intelligence*.
13. Culler, E. A. (1926). 'Studies in psychometric theory.' *J. Exp. Psych.*, pp. 271 and f.
14. Vernon, P. E. (1940). *Measurement of abilities*.
15. McCall, W. A. (1927). *How to measure in education*.
16. Thurstone, L. L. (1925). 'A method of scaling psychological and educational tests.' *J. Educ. Psych.*, 16, pp. 446-457.

¹ In a letter to the writer, Professor Burt points out that, if $F(x)$ is any continuous distribution function, and $dF = f(x)dx$, we can put $y = \int_{-\infty}^x f(x)dx$; and then $dy = dF$: so that such a

distribution can always be converted to rectangular form. He would thus treat the (cumulative) 'distribution function' as prior to the 'frequency function,' instead of vice versa. As to the conditions quoted from (7), p. 118, he argues that it is the relation (not the individuals) that must satisfy the postulates of transitivity and asymmetry. To avoid assuming numerical magnitudes (lengths, measured abilities, etc.) from the start, he suggests beginning with any set, whose elements can be arranged in sequence, in virtue of such a relation (R say). We can then define the 'interval' between x and y as the class of elements lying between x and y (with respect to R), including x but not y (*Principia Mathematica*, II, p. 233). The sum of two such intervals (i.e., of intervals which are half-open, and closed on the same side) will also be an interval; and with each interval we can now associate a measurable 'length' (e.g., by counting the elements it contains). Thus the 'additive law' can be realized for other psychological qualities besides those of task-difficulty. For more general discussions he refers to the 'French School,' esp. Borel, E. (ed.), *Traité du calcul des probabilités* (1924-38), Lebesgue, H., *Leçons sur l'intégration* (1928), Lévy, P., *Théorie de l'addition de variables aléatoires* (1937).

FACTOR ANALYSIS OF ASSESSMENTS FOR ARMY RECRUITS

By CHARLOTTE BANKS

Department of Psychology, University College, London

I. *The Application of Factorial Techniques to Problems of Selection.* II. *Tests and Subjects.* III. *Results for Men :* (i) *Bipolar Analysis ;* (ii) *Group Factor Analysis ;* (iii) *Analysis by Subdivided Factors.* IV. *Results for Women :* (i) *Bipolar Analysis ;* (ii) *Group Factor Analysis ;* (iii) *Analysis by Subdivided Factors.* V. *General Comparisons.* VI. *Summary and Conclusions.*

I. THE APPLICATION OF FACTORIAL TECHNIQUES TO PROBLEMS OF SELECTION

Educational and Occupational Classification of Children. During the thirty years that preceded the outbreak of the recent war a large amount of factorial work was carried out in this country ; but for the most part these investigations were limited to the field of child psychology. Their initial aim was to assist in the classification of school children and of school leavers for purposes of educational or vocational guidance ; and this inevitably led to the problem of classifying cognitive abilities and their assessments. The conclusions reached were based almost entirely on scholastic or comparatively simple psychological tests ; and referred primarily to boys and girls of school age. But the results obtained often threw a suggestive light on the nature of mental capacities and upon the apparent structure of the human mind.

The value of factor analysis as an aid to vocational guidance and personnel selection has sometimes been questioned.¹ But the criticisms seem largely to rest on too narrow a view of the factorial method. Factor analysis, as has often been emphasized,² is merely a special branch of the general technique of multiple correlation. In the ordinary form of the multiple correlation procedure, test-performances and similar assessments are validated or analysed by the aid of *external* criteria ; in factor analysis they are validated or analysed by reference to *internal* criteria. But in almost every new sphere of enquiry *both* types of evidence are desirable, and in the earliest factorial investigations were used side by side.

This double approach was begun by Burt in 1909 ; and was systematically applied, with the aid of teachers and other collaborators, during his work first in the London schools and later as head of the Vocational Branch of the National Institute of Industrial Psychology. Standard tests and rating-scales were compiled both for educational guidance and for occupational guidance and selection. Simple statistical techniques were worked out, with a view to analysing such assessments into a limited number of independent factors—'general,' 'bipolar,' or 'group.'³

In almost all these investigations a large 'general factor' was discovered, roughly identifiable with 'general cognitive efficiency' ; and, wherever sufficiently large samples were obtainable, a small number of 'group factors' were also established. In the earlier investigations the group factors most frequently noted were (i) a Verbal or Linguistic factor, usually dividing into two main subfactors, viz. (a) a factor for isolated words (a 'verbal' factor in the more literal sense), and (b) a factor for the use of consecutive language (a 'literary' factor) ; (ii) a Number factor, again usually splitting into two subfactors for (a) mechanical arithmetic and (b) problem work respectively ; (iii) a

¹ Cf. the discussion between Thomson, G. H. ('The Factorial Analysis of Human Abilities') and Thouless, R. H. ('A Reply'), *The Human Factor*, IX, 1935, pp. 180-5, 358-63.

² Burt, C., 'Principles of Vocational Guidance,' *Brit. J. Psych.*, XIV, 1924, pp. 344f. ; also *Factors of the Mind*, pp. 60-65.

³ *The Factors of the Mind*, pp. 139f., 447f., and refs.

distinctive Practical or visuo-kinæsthetic factor, dividing into such subfactors as (a) a motor factor entering into tests of manual skill, (b) a visuo-spatial factor, entering into tests requiring a comprehension of relations and movements in space, and often (c) a mechanical factor, depending on aptitude for, and often acquired knowledge or experience in, mechanical operations of various kinds. The result was the common suggestion that, with children and young persons at any rate, the most useful battery of tests was one which endeavoured to combine these four main factors—*g*, *v*, *n*, and *k* as they were sometimes briefly labelled.¹

This led to a working conception of the mind as consisting of a hierarchy of factors, differing in complexity or 'level'—a single comprehensive 'general factor' covering all kinds of cognitive abilities, a few broad 'group factors' for the more prominent types of mental processes or contents, and a number of narrower and more specialized 'group factors' differentiated out of these. At the same time, it was emphasized that the "three main types of factor differ only in degree: the 'general factor' is simply the group factor that is most widespread; the 'specific factors' are simply group factors that are most narrowly limited in their operation" (8, p. 19).

To this 'three factor theory' (as it was sometimes not very appropriately termed) the two main alternatives were Spearman's 'unifocal' theory, which attempted to dispense with group factors, and the 'multifocal' theories of Thorndike and (later) Kelley and Thurstone, which attempted to dispense with the general factor. For practical purposes at any rate, the older theory, which recognizes all three types of factor, appears at last to have won fairly wide acceptance, at any rate among educational and vocational psychologists in this country.²

Occupational Classification in the Army. When the Directorate of Selection of Personnel was established at the War Office in 1941, one of its first tasks was "to undertake an analysis of work in the Army, with a view to classifying (if possible) the immense variety of activities into about six or eight main categories, and to validate and standardize suitable tests with a view to classifying new recruits under appropriate occupational categories or 'training recommendations' according to the general ability, special aptitudes, relevant knowledge or skill of each one: for the purposes of these preparatory researches, it appeared that the methods worked out by educational psychologists for analysing and validating mental and scholastic tests for use in schools could be adapted for the special problems of the war."³ As a working hypothesis, it was assumed that the same types of factor and the same broad lines of classification could be expected among adults as had already been found among children, and that the best way of confirming these assumptions would be to apply the same factorial techniques as had proved effective in work upon school pupils.⁴

¹ In addition other special abilities or disabilities were subsequently established, e.g., for immediate and delayed Memory, for various types of Imagery, which do not concern us here. Cf. Burt, C., *Distribution and Relations of Educational Abilities*, 1917, and later L.C.C. Reports; also 'The Relations of Educational Abilities,' *Brit. J. Psych.*, IX, 1939, pp. 55f. A systematic survey of the chief factors revealed by factorial methods will be found in Burt, *The Measurement of Mental Capacities*, Oliver and Boyd, 1926; revised accounts will be found in (10), sect. II, 'Results'; see also *id.*, 'The Abilities of the Mind: a Review of the Results of Factor Analysis,' *Brit. J. Educ. Psych.*, XIX, pp. 70f.

² The combination of one general and several group factors (including a verbal and a practical factor or its equivalents) appears to have been assumed by most of the contributors to the recent symposium on 'The Selection of Pupils for Different Types of Secondary Schools,' *Brit. J. Educ. Psych.*, XVII, XVIII, 1947-8. Even those departments that have favoured other methods of factor analysis than Burt's appear now to have accepted much the same general views: thus, although Dr. Emmett prefers Dr. Lawley's method of factorization and Professor Thurstone's graphical method of rotation his final tables show a set of factors which include *both* a general (or 'basic') factor *and* several group factors: (this *Journal*, II, pp. 3-16).

³ Burt, C., 'British Psychology in War Time,' *Educational Forum*, IX, 1945. Cf. *The Report of the Expert Committee* (9). As statistical member of the War Office Advisory Committee of Psychologists, Professor Burt was asked to suggest suitable statistical procedures; and, during the first few months of the new Directorate's activities, batches of data were referred to him from time to time, and analysed with the assistance of research-workers in his department.

⁴ 'Work for the Services at the Department of Psychology, University College, London.' P.P.(S.C.), 43: quoted in *The Report of the Expert Committee*, *loc. cit. sup.*, p. 77.

Since the following paper is concerned exclusively with factorial results, it should perhaps be emphasized that factor analysis formed only one part of the experimental and statistical enquiries undertaken for the Directorate, and that the study of cognitive and practical abilities constituted only a section of the total field explored. The objects of such analyses were much the same as in the educational sphere: first, to discover what are the basic 'factors' (i.e., the main types or groups) into which human performances can conveniently be classified, and in particular to determine what factors were being measured by various batteries of tests with this or that portion of the adult population; secondly, to analyse new tests in terms of the factors which had already been established, or had been indicated by the needs of job analysis; thirdly, to determine what tests might most usefully be applied, and how the several tests should be summed or weighted, in order to obtain assessments of the various characteristics or aptitudes required for this or that type of training or for this or that type of army occupation.¹

The present paper is concerned with a factorial study of the first large batch of data obtained from psychological testing in the Services. Previously the few investigations upon adults had been limited either to highly-selected batteries of tests or to highly-selected classes of persons, usually small groups of University students.² The men and women for whom the present set of assessments were obtained together numbered well over a thousand, and provide a fairly comprehensive sample of the adult population of the country. The primary outcome of the work was to demonstrate, first, that the methods of factor analysis developed in educational research were also applicable to a far wider range of problems, such as those of personnel selection for civil and military purposes; and secondly, that the chief factors found among children also reappear, with but little modification, among adults.

Acknowledgements. The data and chief conclusions here described were embodied in a thesis on 'Factor Analysis applied to Current Psychological Problems' (1). In the same enquiry similar factorial methods were applied to studies of (a) children and (b) students. These suggest further points of comparison which the reader will find more fully discussed in the thesis itself. I have gratefully to acknowledge permission to use the data for purposes of my original research, and to publish the results in this paper; and am especially indebted to Colonel B. Ungerson (Chief Psychologist to the Director of Manpower Planning) for his kindness in reading my manuscript and for his many helpful comments and suggestions. The responsibility for any errors either of fact or of opinion is of course entirely mine.

II. TESTS AND SUBJECTS

Tests or Traits. The following are the tests with which the present analysis is concerned (cf. 9, pp. 27, 30). The majority had already been in use before the war for purposes of occupational guidance or selection at the National Institute of Industrial Psychology. A few were new inventions devised by members of D.S.P. staff.

1. *Matrix.* The so-called 'progressive matrices test,' consisting of non-verbal analogies of the type introduced by Burt and developed by Raven at University College (20 mins.).
2. *Squares.* The N.I.I.P. test of spatial judgment (10 mins.).
3. *Assembly.* A mechanical assembly test of the Stenquist type (23 mins.).
4. *Bennett.* A modification of the Bennett test of mechanical comprehension (15 mins.).
5. *Agility.* A measure of the time taken to transfer metal rings from two upright pegs to two others 20 feet away (time usually varying from 1 to 2 mins.).

¹ A general summary of the various statistical devices eventually employed, and some instructive comments on their uses and limitations, will be found in Dr. P. E. Vernon's paper (7): see especially his section on 'Factor Analysis,' pp. 144-5.

² With adults the largest investigation of this type previously reported would appear to be that based on data obtained from 2,000 ex-service candidates for the Civil Service (summarized in Burt, C., 'Mental Differences between Individuals,' *Brit. Ass. Ann. Rep.*, 1923, pp. 215-39). However, as the tests then applied were all designed to test general intelligence, not special aptitudes, the factorial results were necessarily limited.

6. *Morse Aptitude*. An American Army test intended to assess aptitude for learning the Morse code. The examinees listen through headphones to 78 pairs of sounds, resembling Morse signals, and are required to say whether the two sounds are the same or different (10 mins. In practice it was found difficult to ensure that all headphones were equally efficient; and, since for this and other reasons it proved to have a low reliability, it was later dropped).

7. *Messages*. A verbal test devised by Nigel Balchin, largely to assess the examinee's ability to interpret messages which have got confused in transit (15 mins. It was found difficult to score the interpretations; and moreover, the factorial results showed that it added little or no information to that already furnished by the other verbal tests. It was therefore subsequently dropped).

8. *Clerical Test*. A verbal test of the instructions type, consisting of five clerical operations: checking names and addresses, filing, coding, and classifying (15 mins.).

9. *Spelling*. Underlining the correct spellings from among six different versions—one correct and five incorrect (5 mins.).

10. *Arithmetic I*. A test of the four arithmetical rules (6 mins.).

11. *Arithmetic II*. A test of mathematical problems (10 mins.).

The following items are also included in the tables: 12. *Age*. 13. *Height*. 14. *Medical Category*.

15. *Educational Standard* (length and type of schooling).

The reliability coefficients for all the tests were over 0.85, except for Assembly (0.76), Agility (0.58), and Morse Aptitude (0.68).

Subjects Tested. The tests were given at Primary Training Centres to recruits entering the Army. The data obtained from the first 1,000 cases (approximately) were submitted to Sir Cyril Burt (as a member of the War Office Advisory Committee) for study and report, and were analysed at University College under his general direction.

(a) *Men*. The number of male recruits amounted to 578. These all took the General Service Battery of tests. Figures were recorded for every item in the foregoing list except Spelling; the marks for the Arithmetic tests, however, were reported as a single score. Since Age (measured in years from birth) gave negative correlations, the signs of the deviations have been reversed; so that the coefficients printed in the tables for men may be regarded as being based on a measure for Youth.

(b) *Women*. The women numbered 595. All took the standard A.T.S. Battery, including the tests numbered 1, 2, 4, 8, 9, 10, and 11 in the list above. In addition Age, Educational Standard, and Medical Category were also reported.

Judged by the frequency distribution for the Matrix test, the men may be regarded as a fairly representative sample of the males of this country. 430 had received no formal educational after leaving the elementary school; the remainder had had some form of further educational—secondary, technical, or the like. As compulsory service was not introduced for women until 1942, the sample of female recruits may be somewhat selected; nearly 30 per cent. were office workers (cf. 6).

III. RESULTS FOR MEN

Correlations. The majority of the correlations were calculated by the product-moment method. For Age, Matrix, Medical Category, and Educational Standard, however, this procedure was not possible; for these three tetrachoric coefficients were calculated for each pair of items, using different borderlines; and the average inserted in the table. These preliminary figures were computed by the staff of D.S.P. With a sample of 578 the standard error of a zero correlation would be ± 0.042 .

(i) *Bipolar Analysis*. The table of correlations was first factorized by Burt's method of Simple Summation, the reduced self-correlations or 'communalities' being determined by successive approximation. In all five factors were extracted. Table IIIA gives the saturation coefficients for the first four.

The fourth factor is barely significant. Two tests were applied. (i) With the chi-squared test the value obtained for P is less than 0.01; this test, however, is not entirely valid. (ii) Taking the standard errors of the observed coefficients as though they had been computed by the full product-moment procedure, we find that out of the 78 residuals on which the fourth factor is based, six are larger than twice the standard error; by chance we should expect barely four.

General Factor. The first factor contributes 41 per cent. to the total variance. This is roughly the proportion usually obtained when factorizing tests of mental ability. Burt has suggested that the factor should be regarded as estimating 'general military efficiency' so far as this can be assessed by a limited battery of tests and traits like the present, with no assessment of temperamental qualities. Judged by the saturations, the items yielding the best estimates for this factor are the Matrix, Clerical, and Arithmetic tests, closely followed by Messages, Educational Standard, Bennett, and Squares, in that order. The remainder yield figures below 0.50. Though the battery contains an unusually large proportion of non-verbal tests and traits, it is satisfactory to find that items so different as the Matrix, Clerical, and Arithmetic tests all furnish high saturations.

Second Factor. The second factor is bipolar; and contributes 7 per cent. to the total variance. It contrasts (a) what may be called the more intellectual characteristics (particularly those assessed by scholastic and verbal tests) with (b) non-scholastic or non-intellectual characteristics. The fact that the Morse and Matrix tests are classed with the scholastic group may be due partly to their novelty and comparative complexity, and partly to the disproportionate number of spatial and practical tests, and of non-cognitive traits, like Youth and Medical Category. The men who do best at the tests in the 'intellectual' group are largely those who have enjoyed some degree of further education. Hence it is difficult to decide whether a difference may not reflect a difference in educational experience quite as much as in intellectual aptitude.

Third Factor. The third factor contributes only 4 per cent. to the total variance. (a) For the seven items which fall into the 'intellectual' group the saturations are decidedly small. This, as we shall see in a moment, taken with the general pattern of the residuals, suggests that a more accurate interpretation might be reached by using 'subdivided factors.' So far as the figures can be trusted, the sign pattern of this and the following factor appears to contrast (1) the more abstract tests which depend largely on good schooling (Clerical, Arithmetic, and Messages, together with Educational Category itself), and (2) the more concrete tests, such as Morse and Matrix. The positive saturation for Height may be due to the fact that members of those social classes that commonly enjoy a secondary education tend to be slightly taller than the average.

(b) Among the more practical or non-intellectual group of tests (1) positive saturations are allotted to the three tests of mechanical aptitude, Assembly, Squares, and Bennett, and (2) negative to the three physical qualities assessed by Medical Category, Youth, and Agility. Good physique and muscular agility are characteristics which we might expect to be more strongly marked in younger recruits as contrasted with older.

Fourth Factor. The third bipolar factor contributes only 3 per cent. to the total variance. Except for Youth and Height, the classification suggested by the sign pattern is the same as that indicated by the preceding factor. This repetition of the sign pattern (with the second half reversed) is very characteristic of cases in which 'subdivided factors' appear.¹

(ii) *Group Factor Analysis.* In order to obtain an analysis in terms of positive coefficients alone, the correlations shown in Table I were factorized by Burt's Group Factor Method.² The simplest procedure is to take the threefold grouping indicated by the larger saturations for the bipolar factors, and then, with the correlation matrix thus partitioned into nine submatrices, to apply the usual formula. The saturations thus obtained are shown in Table IIIB.

¹ The reason has been stated by Burt. With the ordinary analysis "the first of the two bipolar factors necessarily assumes the existence of significant cross-correlations [between the two main groups]; and inevitably produces non-zero figures in the corresponding part of the hypothetical hierarchy that represents it. When, in point of fact, there are no such significant residuals, the next bipolar factor, with the same lines of division, but a partly reversed sign pattern, has to be invoked to cancel the fictitious cross-correlations introduced by the previous bipolar" (this *Journal*, II, p. 54: see also *ibid.*, Tables IIA and III).

² *Factors of the Mind*, pp. 477f.

TABLE I. CORRELATIONS : MEN

Assessments	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Clerical	(.810)	.773	.714	.752	.293	.453	.737	.539	.521	.313	.173	.245	.241
2. Memory	.773	(.752)	.675	.712	.265	.490	.643	.484	.538	.239	.126	.132	.257
3. Education	.714	.675	(.669)	.731	.255	.404	.560	.466	.514	.310	.250	.186	.267
4. Arithmetic	.752	.712	.731	(.771)	.301	.455	.689	.526	.557	.234	.165	.239	.273
5. Height	.293	.265	.255	.301	(.132)	.162	.260	.224	.265	.123	.052	.069	.002
6. Morse	.453	.490	.404	.455	.162	(.396)	.529	.363	.422	.164	.296	.241	.230
7. Matrix	.737	.643	.560	.689	.260	.529	(.767)	.595	.630	.364	.340	.303	.232
8. Squares	.539	.484	.466	.526	.224	.363	.595	(.579)	.607	.492	.193	.196	.287
9. Bennett	.521	.538	.514	.557	.265	.422	.630	.607	(.633)	.498	.182	.265	.276
10. Assembly	.313	.239	.310	.234	.123	.164	.364	.492	.498	(.578)	.151	.219	.284
11. Youth	.173	.126	.250	.165	.052	.296	.340	.193	.182	.151	(.403)	.208	.287
12. Agility	.245	.132	.186	.239	.069	.241	.303	.196	.265	.219	.208	(.236)	.306
13. Medical Category	.241	.257	.267	.273	.002	.230	.232	.287	.276	.284	.287	.306	(.589)

TABLE II. CORRELATIONS : WOMEN

Assessments	1	2	3	4	5	6	7	8	9	10
1. Clerical	(.998)	.728	.540	.703	.759	.065	.665	.477	.299	.125
2. Spelling	.728	(.939)	.586	.598	.469	.045	.540	.361	.307	.048
3. Education	.540	.586	(.708)	.590	.635	.130	.362	.252	.280	.071
4. Arithmetic I	.703	.598	.590	(.750)	.703	.060	.491	.452	.208	.034
5. Arithmetic II	.759	.469	.635	.703	(.827)	.076	.559	.464	.330	.102
6. Age	.065	.045	.130	.060	.076	(.128)	—	.016	.026	—
7. Matrix	.665	.540	.362	.491	.559	.018	(.685)	.631	.414	.066
8. Squares	.477	.361	.252	.452	.464	.016	.631	(.690)	.344	.025
9. Bennett	.299	.307	.280	.208	.330	.026	.414	.344	(.396)	.088
10. Medical Category	.125	.048	.071	.034	.102	—	.066	.025	.088	(.053)

TABLE III. FACTOR-SATURATIONS : MEN

Assessments	A. Bipolar Factors				B. Group Factors			C. Subdivided Factors					vii
	I	II	III	IV	i	ii	iii	iv	v	vi			
1. Clerical ..	.825	+ .339	+ .021	+ .074	.672	.596	—	—	—	—	—		
2. Messages ..	.763	+ .382	+ .034	+ .043	.587	.650	—	—	—	—	—		
3. Education ..	.753	+ .217	+ .099	+ .161	.658	.411	—	—	—	—	—		
4. Arithmetic ..	.805	+ .325	+ .008	+ .108	.659	.570	—	—	—	—	—		
5. Height ..	.302	+ .129	+ .125	+ .075	.243	.217	—	.216	—	—	—		
6. Morse ..	.579	+ .045	+ .222	+ .096	.567	.159	—	.593	—	—	.221		
7. Matrix ..	.835	+ .072	+ .065	+ .216	.815	.253	—	.673	.177	—	—		
8. Squares ..	.699	+ .134	+ .228	+ .142	.741	—	.163	—	.261	.151	—		
9. Bennett ..	.744	+ .115	+ .194	+ .171	.797	—	.141	—	.103	.245	—		
10. Assembly ..	.496	+ .441	+ .366	+ .006	.459	—	.895	—	.504	.634	—		
11. Youth ..	.351	+ .203	+ .425	+ .154	.311	—	—	—	.277	—	.407		
12. Agility ..	.355	+ .207	+ .202	+ .018	.338	—	—	—	.281	—	.328		
13. Medical Cat.	.436	+ .409	+ .161	+ .444	.379	—	—	—	.339	—	.427		
Contribution to Variance ..	41.2	7.3	4.0	3.1	34.3	10.8	6.5	3.8	4.0	3.7	3.9		

TABLE IV. FACTOR-SATURATIONS : WOMEN

Assessments	A. Bipolar Factors				B. Group Factors			C. Subdivided Factors					vii
	I	II	III	IV	V	i	ii	iii	iv	v	vi	vii	
1. Clerical ..	.897	+.132	+.194	+.179	+.296	.935	.103	—	—	—	—	—	—
2. Spelling ..	.776	+.192	+.286	+.435	+.136	.775	.341	—	—	—	—	—	—
3. Education ..	.696	+.108	+.181	+.290	+.189	.796	—	.324	—	—	—	—	—
4. Arithmetic I ..	.770	+.250	+.127	+.148	+.032	.737	—	.290	—	—	—	—	—
5. Arithmetic II ..	.826	+.322	+.296	+.116	+.088	.620	—	.485	—	—	—	—	—
6. Age ..	.085	+.140	+.130	+.067	+.124	.032	—	.181	—	—	—	—	—
7. Matrix ..	.738	+.339	+.153	+.019	+.027	.668	—	—	—	.281	.312	—	—
8. Squares ..	.624	+.436	+.178	+.090	+.271	.513	—	—	—	.334	.267	—	—
9. Bennett ..	.452	+.331	+.239	+.155	+.077	.385	—	—	—	.294	—	.321	—
10. Medical Cat.	.099	+.039	+.092	+.047	+.168	.098	—	—	—	.007	—	.123	—
Contribution to Variance ..	43.2	6.7	4.0	3.9	2.8	39.0	1.3	4.6	7.4	2.8	1.7	1.2	—

The first factor is the 'basic factor,' with positive saturations for every trait. It accounts for 34 per cent. of the variance. The size of the saturation coefficients places the traits in much the same order as the size of the first factor-saturations obtained with the bipolar analysis. The Bennett and Squares tests, however, are now placed second and third, whereas in the bipolar analysis they were placed only sixth and seventh. The 'basic' factor, obtained by the group factor method, therefore appears to be somewhat biased by tests of spatial and mechanical aptitude.

At first sight it might seem that, whichever method of analysis is adopted, the Matrix test appears the best single test of general efficiency. It would, however, be somewhat rash to conclude (as some have apparently done) that this particular test might serve as a satisfactory and self-sufficient test of adult intelligence. So far as military efficiency is concerned, the figures from these and later batches seem clearly to show that a weighted set of a few selected tests are distinctly more effective than any single test, and that a set which includes verbal tests of intelligence is, on the whole, the most reliable and the most useful.¹

Of the three group factors, the largest consists chiefly of cognitive tests. The highest saturations are furnished by Messages, the Clerical, Arithmetic and Education tests. Height has a small saturation for this factor. The second group factor comprises the three tests of mechanical aptitude, namely, Assembly, Squares and Bennett; it is possibly akin to the so-called *k*-factor. The third group factor is formed by Medical Category, the Agility Test, and Youth. It appears to represent the general physical and muscular fitness characteristic of youth.

(iii) *Analysis by Subdivided Factors.* The foregoing set of non-overlapping group factor-saturations yields a tolerably good fit to the observed correlations. With few exceptions the residuals appear devoid of statistical significance. Nevertheless some are high enough to be at least suggestive of supplementary factors, and fall into patterns which can hardly be due to chance. In particular, the large group of intellectual tests or traits appears to be divisible into two subgroups; while the residual cross-correlations between two smaller non-intellectual subgroups seem to require a broader group factor covering all six.

Accordingly, it was thought instructive to attempt a fresh analysis of the entire correlation table by the method of 'subdivided factors.' Some such procedure, as we have already seen, is further indicated by the residuals obtained in the original bipolar analysis and the sign patterns of the factor-saturations in the bipolar matrix. It must, however, be remembered that the set of tests and traits here analysed was compiled for practical rather than for theoretical purposes. From an analytic standpoint, therefore, it forms a somewhat heterogeneous and ill-balanced collection. Hence, it is not to be expected that the more complex procedure will yield very clear or trustworthy results.

A preliminary analysis into subdivided factors was first carried out by direct calculation, using the ordinary formulæ for group factor analysis. A better fit was then secured by successive approximation. This consists in subtracting from the observed correlations all the expected correlations due to group factors thus calculated, leaving in turn only one of the factors to be recalculated.

The pattern of factors² so obtained is shown in Table IIIc. It will be seen that the whole battery

¹ This emerges clearly in the analysis of general service follow-up results (a point for which I am indebted to Colonel Ungerson). Judged by reports received for the earlier batches after training, the average validity of the tests appears to range from about .20 or less for Agility, .30 for Matrix, to between .35 and .55 for the Clerical and Arithmetic tests. The multiple correlation for a Combination of Messages, Clerical and Arithmetic tests ranges from about .45 to .65, increasing, on correction for selection, to between .55 and .75. But the attempt to assess validity is a highly complex problem, which no doubt will be dealt with more fully by others in the light of more recent and more extensive data. I may add that the high selective value of tests or assessments which appear to depend on educational progress would seem to be attributable, not so much to the effects of educational opportunity as such, but rather to the fact that, when we are dealing with a large and heterogeneous sample of the adult population, educational progress is itself one of the most reliable indexes of intelligence available.

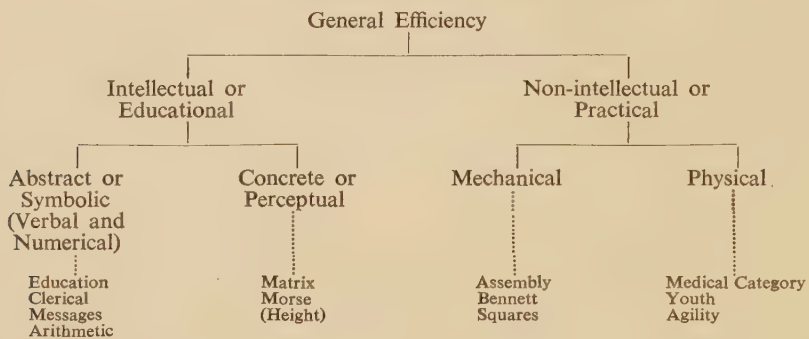
² This type of factor-pattern has frequently appeared in earlier work with mental tests: (a list of instances will be found in Burt's *Notes on Factor Analysis: IV. Results*). It is perhaps most clearly seen in the factorization of bodily measurements, where the errors of measurement do not blur the factor-patterns so much as they do with mental assessments. Cf., for example, Banks (1), Table VI, p. 103, *id.*, *Ann. Eugen.*, XIII, 1947, Table 3, p. 242, and Burt, *Psychometrika*, XII, 1947, Table 4, p. 185.

of thirteen traits is first divided into two main groups—what may be loosely called an intellectual and a practical group. The practical group is then divided into two subgroups along much the same lines as before. Similarly, the intellectual group, consisting for the most part of traits having a fairly high correlation with general mental efficiency, is then divided into two subgroups, which may perhaps be provisionally termed an abstract or symbolic¹ and a concrete or perceptual group respectively.

Matrix shows a small overlap with the 'practical' factor, and Morse Aptitude with the 'physical' factor. This is intelligible: the Matrix test is essentially non-verbal, and so might be in part correlated with other tests of non-verbal efficiency. The Morse Aptitude test depends on acuity of hearing, which would be poorer in recruits of low medical category or more advanced age. With this final method of analysis, all the residuals, with one exception, prove to be less than ± 0.050 . The exception is the somewhat large residual correlation relating Matrix to Youth: this agrees with the findings of several other investigators who have noted that efficiency with this test shows a specific decline with increasing age.

The hierarchical relations of the several tests and traits can be seen most clearly if factors and subfactors are arranged in the form of a genealogical tree² (cf. Fig. 1).

FIG. 1. SCHEME OF FACTORS: MEN



IV. RESULTS FOR WOMEN

The correlations were calculated in the same way as before; and are set out in Table II.

(i) *Bipolar Analysis.* As with the men, the correlation table was first factorized by Simple Summation; and, as before, five factors were extracted. Judged by the same criteria, the fourth factor now appears fully significant at the 5 per cent. level. None of the residuals on which the fifth factor is based reach a figure that is twice as large as the standard error of the original correlation. The saturations are shown in Table IVA.

General Factor. As in the case of the men, the general factor is perhaps best described as general military efficiency (in so far as this is measured by the battery). The first factor contributes 43 per cent. to the total variance—rather more than with the men. This agrees with the difference found in other investigations, whether with children, students, or the general population: in nearly all, the abilities and attainments of the males tend to appear more specialized than those of the women or girls.³ The items having the highest saturations are the Clerical test, Educational Category, Spelling,

¹ The term 'symbolic' is used to indicate the fact that most of the tests in this group involve thinking in terms, not of concrete visual or auditory percepts, but of abstract symbols—verbal, numerical, or (in the case of the clerical test) code-labels.

² Figures 1 and 2 are taken from the 'factorial trees' given as Tables IV and VIII in my thesis (1, 1945, facing pp. 5 and 22).

³ Cf. for example, Burt, C., and Moore, R. C., *Mental Differences between the Sexes*, J. Exp. Ped., I, 1912, pp. 362f.

Arithmetic I, and Matrix—all with coefficients over 0.70. With the women Medical Category and Age have virtually zero saturations, and thus do not fall within the same general class as the traits measured by the psychological tests. Thus with this battery and sample, the first factor, which we should expect to be a 'general factor,' proves to be really a broad group factor. This suggests that, with the women, it may be more nearly akin to what is commonly termed 'general intelligence.'

It may be noted that, whereas with the male recruits the Matrix test furnished the highest saturation for the first factor in both bipolar and group analyses, with the women its saturation ranks only fifth. Several explanations may be suggested. First, owing to the very different types of work which men and women are required to perform in the army, the most suitable items on which training-recommendations can be based are somewhat different for the two sexes. Thus, as we have already seen, the men's battery contains a large number of non-scholastic and non-verbal tests and traits: only three out of the thirteen items are definitely educational; with the women, on the other hand, half the items fall into the scholastic class, and only three are definitely practical. Hence with them the assessments for the general factor will necessarily have a more scholastic or verbal bias. As a result the Matrix test, which is essentially non-verbal, might be expected to show a lower correlation with the general factor. Nevertheless, this can hardly account for all the difference. In a later set of data, where the battery used for the women was almost identical with that for the men, and where the sample tested was virtually unselected, the Matrix test still showed a comparatively low saturation with the women; and the conclusion was drawn that "non-verbal tests, including Matrix, are less adequate as tests of general efficiency with women than with men."¹ These further results, therefore, suggest that we may here be partly concerned with a fundamental difference between the two sexes, general efficiency in women apparently tending either to take a more verbal form or else to express itself more readily in verbal terms.

Second Factor. The first bipolar factor contributes 7 per cent. to the total variance. As with the men, this factor appears to contrast (a) what we have called the more intellectual characteristics with (b) the non-intellectual characteristics. With two items, however, Matrix and Age, the classification is altered. With the women Matrix obtains a fairly high positive saturation, whereas with the men it had a negligible negative saturation. As a result it is now grouped with the more practical tests. The change is probably connected with the two main differences just discussed. First, the difference between the tests included in the two batteries is almost bound to alter the resulting bipolar factors.² Secondly, owing to the verbal bias of the women, it seems possible that with them the perception of relations of space and form depends on a more highly specialized aptitude than with the men.³

A further difference between the two sexes is that, whereas with the men Age had a negative saturation for this factor, and Youth therefore a positive, with the women we find that Age has a positive saturation. With this group the most obvious explanation is that, at this stage of enlistment, there was a tendency for older recruits for the A.T.S. to be of better educational standard and to show greater scholastic aptitude than the younger.⁴ At this period a large proportion of the older recruits were clerks and office workers; the younger women came largely from the domestic servant class; factory workers usually opted to do their war service in the factories (6).

¹ From a Report by Senior Commander Wickham: I am much indebted to Colonel Ungerson for permission to quote this report. It may be added that, with the women more particularly, the saturations both for the general and for the supplementary factors change somewhat in the pattern they present, when figures for different intellectual levels are analysed separately. This is merely one of the many instances in which evidence was found that in different groups the same tests might elicit somewhat different mental processes or abilities. It must therefore be emphasized that the results reported above provide merely a first broad indication of average or general tendencies.

² Burt has shown by a simple example (quoted in (1)) how a change of balance in constructing a test-battery may easily produce a change of sign in one of the smaller bipolar factor-saturations. Thus, if a composite battery has twice as many verbal as non-verbal tests, the summation method requires that one or more of the verbal tests that depend least of all on verbal facility shall be transferred to the non-verbal group: otherwise the residuals on which the bipolar factor is based would not add up to zero. Here the change from a battery which was overweighted with practical tests (see above, p. 80) to one in which the number of verbal tests has been increased has probably caused the Matrix test to change over from the verbal to the non-verbal group.

³ This again seems confirmed by the results since obtained by Senior Commander Wickham. With her sample of the women the Matrix test has a positive saturation for the first bipolar factor, and, after rotation, shows a fairly large saturation for the group factor labelled "*k : g*."

⁴ It does not follow that this is the only cause. With later intakes, where there was less difference in the occupations of the women at different age-levels, there was still some evidence that spatial, mechanical, and non-verbal abilities generally tend to decline with age, while verbal and educational, if anything, improve.

Factor Analysis of Assessments for Army Recruits

Third Factor. The third and fourth factors each contributes only 4 per cent. to the total variance. (a). Among the intellectual or scholastic group of tests the second bipolar factor appears to separate (1) the more elementary tests, which depend largely on mechanical memory and accuracy (Spelling, Clerical and Arithmetic I), from (2) the harder tests (Arithmetic II, Age, and higher Educational Category). (b). Among the non-intellectual tests or traits, this factor contrasts (1) Squares and Matrix with (2) Bennett and good Medical Category. Since the saturation for Medical Category is small, it might be fair to say that the main contrast here lies between the concrete perception of space, on the one hand, and a more abstract knowledge of elementary mechanical principles¹ on the other. However, we might also expect that women who had been interested in mechanical subjects would be found largely among those of more robust health; whereas those with poor vision (an important item in assessing medical category in the Army) might be handicapped in tests like Squares and Matrix which require quick and accurate discrimination of visual patterns.

Fourth Factor. For most of the tests and traits the fourth factor yields low saturations having the same sign as those furnished by the third factor. Education and Spelling, however, have comparatively large saturations, with signs opposite to those for the third factor. Irregular saturations of this kind are a common phenomenon in small batteries, where one or two of the traits have marked specific factors. Apart from this effort to express specific as common factors, the fourth factor appears merely to reinforce the third.

Fifth Factor. The fifth factor contributes only 3 per cent. to the total variance, and is, as we have seen, below the borderline for significance. Its signs repeat the classification of the third factor, and are reversed for the last two groups. The only exception is Matrix, for which the saturation is nearly zero. This, as already noted, is a frequent result where the factors ought to be separately subdivided, but where, owing to the requirements of the ordinary summational procedure, the later bipolar factors are made to extend over the whole set of traits instead of being limited to the two separate sections.

(ii) *Group Factor Analysis.* An analysis into non-overlapping group factors was next carried out by the same method as was used for the men. The saturations are given in Table IVb.

In determining the grouping of the traits the size of the bipolar saturations as well as the signs were taken into account. Thus the large negative saturations shown by the two Arithmetic tests in the first bipolar factor (Factor II) indicated that they should be kept together to form (with Age and Education) the first of the three small groups (Factor ii); the large positive saturations shown by Squares, Matrix and Bennett for the first bipolar factor indicated that they would form the second group (Factor iii); the Spelling and Clerical tests are kept together by the big saturations in Factor III: since the only large saturation for Medical Category appears in Factor V, this item was grouped with the Spelling and Clerical tests which also have fairly large saturations of the same sign for this factor. The outcome is, as before, to partition the entire correlation table into $3 \times 3 = 9$ submatrices, to which the usual formula for three non-overlapping group factors can readily be applied.

The saturations for the first or 'basic' factor show a wider range than those obtained by the bipolar analysis. The order, however, remains much the same (correlation, 0.81). In the first group factor Arithmetic II has the highest saturation, and Age (as might be expected) the lowest. In the second group factor Matrix has the highest saturation, and Squares and Bennett the lowest. The third group factor would seem to be primarily a verbal factor. No separate factor of this type was discernible among the men, because no test of Spelling was included; and hence with them the Clerical test, which is also verbal, was merged into the other scholastic tests. The positive saturation for Medical Category is not altogether unexpected: women of a literary or secondary (grammar) school education are on the average healthier and better developed physically.

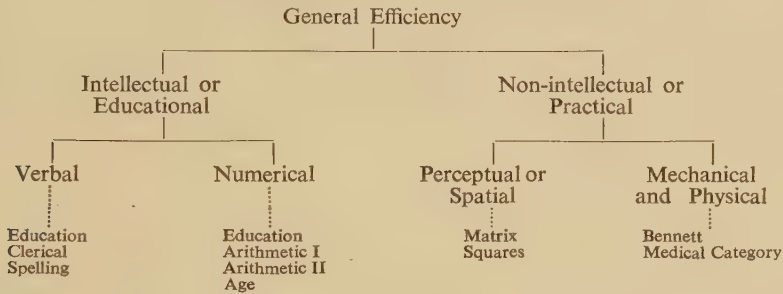
(iii) *Analysis by Subdivided Factors.* As in the case of the men, the simple and straightforward analysis into three non-overlapping group factors leaves small but suggestive residuals. The arrangement of these residuals suggests that a better fit might be obtained by a scheme of subdivided factors, with broad group factors overlapping the narrower. Accordingly, this procedure was applied to the correlation table for women along the same lines as before.

The results are shown in Table IVc. As with the men, there is first a broad division into the more intellectual traits on the one hand and the more practical traits

¹ It will be noted that the Bennett test has particularly low correlations and saturations with the women. This seems to indicate that with different groups the same test may elicit very different qualities.

on the other. Each of these in its turn then splits into two subgroups. The intellectual group splits into (a) a verbal subgroup and (b) an arithmetical or numerical subgroup ; the non-intellectual into (a) a perceptual or spatial subgroup (Squares and Matrix) and (b) mechanical and physical subgroup (Bennett and good Medical Category). Once again, however, it must be insisted that the whole battery is so small and so heterogeneous that the results cannot be regarded as conclusive. The final classification of the tests and traits, indicated by the factorial analysis, is shown diagrammatically in Fig. 2.

FIG. 2. SCHEME OF FACTORS : WOMEN



V. GENERAL COMPARISON

In the main, the results obtained with these adult men and women seem to be consistent with those previously obtained from children of school age. With both men and women the resulting factor-pattern shows, first a general or basic factor covering practically all the traits, and secondly, two broad group factors, which show a clear tendency to split into smaller subfactors. The general factor accounts for a larger proportion of the variance than all the supplementary factors put together. It may be added that this general type of factor-pattern appears to have been fully confirmed by analyses of more recent correlation tables (hitherto unpublished), which have since been obtained from the fighting services.

The discovery of a comprehensive general factor, in adults as in children, is perhaps the most striking and significant result of all. It is the more noteworthy, because, until quite lately, the trend of opinion in this country has been moving away from the emphasis formerly laid upon the general factor, particularly by the adherents of the 'single factor theory.' In the United States there has always been a strong disposition to deny the need for any such factor: even those who found such a factor among children have been inclined to interpret it, not as confirming the notion of an innate general ability, like 'intelligence,' but rather as expressing a difference in rate of maturation during the years of growth—a purely temporary difference which should virtually disappear when the individuals tested have passed beyond the years of immaturity and become fully adult. In this country Burt has maintained that the general cognitive factor, when appropriately measured among school children, is largely, if not mainly, innate. Nevertheless, he has shown that among younger children its influence is greater and more conspicuous, and that, with increasing age, its contribution to individual variability steadily diminishes.¹ On these and other grounds, many psychologists at the outbreak of war believed that among the adult population the importance of the general factor might prove to be almost negligible. If the foregoing results can be trusted, however, it would seem clear that, even with full-grown men and women, the existence of the general factor may be safely assumed, and that its influence is not much smaller with them than with older pupils still at school.

The fact that the more specialized factors make so small a contribution to the total variance may seem at first sight surprising. With both sexes the first bipolar factor contributes only 7 per cent.—a proportion equal to that found by Burt among eight-year-old children, and much less than that found

¹ *Mental and Scholastic Tests*, 1921, p. 266 ; *Brit. J. Educ. Psych.*, XIII, 1943, p. 132 and refs.

by him among children of 12 and 13. In my own investigations (1, pp. 163f. and 189f.) the first bipolar factor was responsible for as much as 14 per cent. with data obtained from older children with the Stanford Binet tests, and about 12 per cent. with assessments obtained from College students. With the Army data the low figure obtained seems attributable, not to the age or the selection of persons, but rather to the way the tests and traits have been selected. For practical reasons it was obviously desirable that, so far as possible, any overlapping among the tests should be avoided ; and many of the traits—Height, Age, and Medical Category, for example—are only remotely related to mental efficiency as measured by the tests in common use with school children.

One final point emerged as a result of the foregoing analysis. Among those who have accepted the more eclectic view that both a general and a number of group factors must be recognized, many believed (and some still believe) that it should be possible to devise a separate test for each factor—a test like Matrix to measure *g* and others to measure special aptitudes, such as *v*, *n*, or *k*. With the men, and still more with the women, it is impossible to maintain that a single test, like Matrix, could serve as a satisfactory test of general efficiency. Here, as elsewhere, the best multiple correlations will obviously be obtained by using a weighted sum based on a composite battery which includes tests of widely different kinds. Even for the more special aptitudes variously weighted sums of two or more tests are far more effective than marks taken from isolated tests constructed to measure distinct abilities on *a priori* grounds.

VI. SUMMARY AND CONCLUSIONS

1. A factorial analysis has been carried out for the correlations between personnel assessments obtained for 1,173 Army recruits (578 men and 595 women). With so large a sample of adults, the data provide material for studying adult abilities on a more extensive scale than anything hitherto available to psychologists in this country.

2. With both men and women a preliminary analysis by simple summation yields at least three factors that are fully significant—one general and two bipolar. These can readily be expressed in terms of a basic factor supplemented by three non-overlapping group factors. However, the patterns of residuals obtained with both forms of analysis suggest that a better interpretation can be secured by using subdivided group factors, namely, two broad group factors, each including about half the traits, and each covering a pair of narrower group factors.

3. With all the methods of analysis, there is among both men and women a marked 'general' or 'basic' factor, underlying all the assessments, and accounting for about 40 per cent. of the variance. It may be regarded as a factor of general efficiency ; and is thus analogous to, but not identical with, general intelligence.

4. The first bipolar factor divides the assessments into two main groups or classes—first, intellectual or educational, and secondly, non-intellectual or practical. With the men the former separates into an abstract or academic subgroup (including both verbal and arithmetical abilities) and a concrete or perceptual subgroup ; the latter into a mechanical subgroup and a subgroup depending chiefly on physical and muscular fitness. With the women, owing partly to a difference in choice of tests, the former group separates into a verbal and an arithmetical subgroup ; the latter into a spatial and a physical or mechanical subgroup.

5. The general organization of mental abilities appears to be much the same for both men and women. There are, however, minor differences in the detailed results. First, there is evidence for a somewhat stronger verbal bias in the factor of general efficiency among the women : owing to its non-verbal form, the Matrix test therefore proves to be less appropriate for testing their general efficiency. Secondly, owing to their smaller aptitude or experience in mechanical and arithmetical work, tests involving processes of this kind seem to depend on more specific factors among the women than among the men.

6. The scheme of tests and assessments was planned primarily for practical purposes, and cannot be regarded as furnishing a complete or representative sample of mental abilities. But on the whole, so far as the data go, the analysis indicates a hierarchy of subdivided factors, which appears to be of much the same general type for both adult men and women as for children of school age. The chief differences would seem to be that with adults (i) the general and first bipolar factors account for rather less of the total variance, and (ii) the first and main classification separates those with an intellectual bias or training from those with a more practical, whereas with children it usually separates those with a verbal bias from those with a non-verbal.

REFERENCES

1. Banks, C. (1945). *Factor-Analysis applied to Current Psychological Problems, with Special Reference to Data from H.M. Forces.* Ph.D. Thesis, University of London Library.
2. Burt, C. (1942). 'Psychology in War.' *Occup. Psych.*, XVI, 95-110.
3. Burt, C. (1943). 'Validating Tests for Personnel Selection.' *Brit. J. Psych.*, XXXIV, 1-19.
4. Burt, C. (1944). 'British Psychology in Wartime.' *Presidential Address, British Psychological Society.*
5. Burt, C. (1944). 'Statistical Problems in the Evaluation of Army Tests.' *Psychometrika*, IX, 219-235.
6. Mercer, E. O. (1945). 'Psychological Methods of Personnel Selection in a Women's Service.' *Occup. Psych.*, XIX, 180-200.
7. Vernon, P. E. (1947). 'Statistical Methods in the Selection of Navy and Army Personnel.' *J. Roy. Stat. Soc. Sup.*, VIII, 139-153.
8. Consultative Committee of the Board of Education (1924). *Report on Psychological Tests of Educable Capacity.* H.M. Stationery Office.
9. Expert Committee (1947). *Report on the Work of Psychologists and Psychiatrists in the Services.* H.M. Stationery Office.
10. Burt, C. (1947). 'Factor Analysis : Aims and Chief Results,' *ap. Miscellanea Psychologica : Albert Michotte.* Louvain : Editions de l'Inst. Sup. de Philosophie.

FACTOR ANALYSIS BY LAWLEY'S METHOD OF MAXIMUM LIKELIHOOD

By W. G. EMMETT

Education Department, Moray House, University of Edinburgh

I. *Introduction.* II. *Working Methods.* III. *Comments.* IV. *The Standard Error of Factor Loadings.* V. *The Standard Error of a Residual.* VI. *Summary.*

I. INTRODUCTION

In a recent enquiry the writer (1) used D. N. Lawley's ¹ Maximum Likelihood Method of Factor Analysis (2, 3, 4) to establish the significance of a spatial factor which a centroid analysis had failed to detect. This was possible in that Lawley's method provides a satisfactory test of the significance of the residual matrix ; tests of significance of residuals obtained by other methods of analysis are only crude and may lead to erroneous conclusions. Lawley's method of analysis has been little used, perhaps on account of the laborious calculations involved ; the labour, however, is well spent and is usually small compared with that of collecting test scores and computing correlations.

Lawley's description of his method (4) relates to an analysis into two factors, and requires amplification when more than two factors are in question. We therefore give here, with Dr. Lawley's assent, a scheme of analysis for three or more factors, showing arithmetical checks that may be applied and how time may be saved. Here and there our subject matter may overlap that given elsewhere by Dr. Lawley ; this has been done for the sake of continuity and completeness.

II. WORKING METHODS

For purposes of illustration the intercorrelations of nine variables, obtained by pooling Slater's (5) 11 + correlations described in the above article (1), are taken as starting point. The original scores were derived from 211 children. Table I gives the correlation matrix, *R*, with unity in the diagonal cells.

TABLE I. THE MATRIX OF CORRELATIONS, *R*

Test	1	2	3	4	5	6	7	8	9
1	1·0000	·5232	·3950	·4706	·3455	·4262	·5761	·4338	·6393
2	·5232	1·0000	·4792	·5060	·4181	·4619	·5469	·2829	·6445
3	·3950	·4792	1·0000	·3554	·2701	·2536	·4524	·2185	·5038
4	·4706	·5060	·3554	1·0000	·6906	·7909	·4427	·2852	·5050
5	·3455	·4181	·2701	·6906	1·0000	·6794	·3825	·1488	·4091
6	·4262	·4619	·2536	·7909	·6794	1·0000	·3721	·3138	·4721
7	·5761	·5469	·4524	·4427	·3825	·3721	1·0000	·3846	·6801
8	·4338	·2829	·2185	·2852	·1488	·3138	·3846	1·0000	·4700
9	·6393	·6445	·5038	·5050	·4091	·4721	·6801	·4700	1·0000
Sum	4·8097	4·8627	3·9280	5·0464	4·3441	4·7700	4·8374	3·5376	5·3239

¹ The writer is greatly in the debt of Dr. D. N. Lawley for his help over difficult aspects of the present subject, for reading the manuscript of this article, and for many useful suggestions.

A centroid analysis, using the highest correlation in each column as communality, is first made. The results show that two factors are almost certainly significant, whilst a third remains in doubt. The centroid loadings are taken as first approximations; any guessed loadings within reason may be used, but the more accurate the initial loadings, the fewer the iterations required. We start the working with two factors only; when constancy of loadings has been achieved, the residuals are tested for significance, and if significant, further iterations with three factors are carried out. The procedure in the first iteration is now described, the working being given in Table II.

1. Rows l_1 and l_2 in Block A of Table II are the centroid loadings in Factors I and II, the sum of each row being given in the check column. Row s^2 is the variance of each test that remains after removal of the variance due to the two factors; it may be provisionally called the specific variance. Thus for test 1, $\cdot 4304 = 1 - (\cdot 7184)^2 - (\cdot 2312)^2$. Block B of the table is concerned with finding the second approximation to the loadings of Factor I.

2. Row a_1 , the subscript denoting the factor which is being evaluated, is obtained by dividing l_1 by the corresponding value of s^2 for each test. Thus for test 1, $1\cdot 6691 = \cdot 7184/\cdot 4304$.

3. The elements of row b_1 are the inner products of row a_1 with the successive columns or rows of the original correlation matrix, R . Thus for test 1, $10\cdot 0496 = 1\cdot 6691 \times 1\cdot 0000 + 1\cdot 6024 \times \cdot 5232 + \cdot 8603 \times \cdot 3950 + \dots + 3\cdot 1773 \times \cdot 6393$. It is convenient to write row a_1 on a strip of paper in register with the columns of R and to place the strip immediately above the row in R of which the product is being evaluated. The entries in row b_1 are checked by noting that their sum is equal to the inner product of row a_1 and the column sums of the correlation matrix.

4. Row c_1 is formed by subtracting row l_1 from row b_1 . This and other later additions and/or subtractions of rows are facilitated by using strips of paper with slots cut at the appropriate places, so that only the figures which take part in the computation are exposed to view. The sum of row c_1 is entered in the check column, and checked against the difference between the sum of row b_1 and the sum of row l_1 ; thus $83\cdot 8330 = 90\cdot 0253 - 6\cdot 1923$.

5. The quantity k_1^2 is obtained by forming the inner product of row a_1 and row c_1 ; thence $1/k_1$ is found.

6. Row l_1' is formed by multiplying the elements of row c_1 by $1/k_1$: it is checked by noting that the sum of the row equals $1/k_1$ times the sum of row c_1 , i.e., $6\cdot 1146 = \cdot 072937 \times 83\cdot 8330$. The terms of row l_1' give the second approximations to the loadings in Factor I.

7. The second approximations to the loadings in Factor II are now computed in Block C. Row a_2 is obtained by dividing l_2 by the corresponding value of s^2 for each test. Thus for test 1, $\cdot 5372 = \cdot 2312/\cdot 4304$.

8. The terms of row b_2 are the inner products of row a_2 with the successive columns or rows of the original correlation matrix, R . The computation is performed as for row b_1 in Table II. Care in observing the correct signs is here necessary. It is useful to enter the negative quantities in red ink on the strip of paper suggested in paragraph 3. Thus far the procedure for Factor II is the same as for Factor I. An extra stage now enters.

9. The quantity i_2 is the inner product of rows a_2 and l_1' , and row c_2 is the product of row l_1' and i_2 . The sum of row c_2 checks with the product of the sum of row l_1' and i_2 . The row c_2 may be regarded as a correction to eliminate the effect of Factor I from row b_2 .

10. Row d_2 is obtained by subtracting rows l_2 and c_2 from row b_2 , checking that $2\cdot 1779 = -9\cdot 6700 - \cdot 0121 - (-11\cdot 8600)$.

11. The inner product of rows a_2 and d_2 gives k_2^2 , whence $1/k_2$ is found.

12. Finally the product of row d_2 and $1/k_2$ gives row l_2' , which is the second approximation to the Factor II loadings. The sum of this row, $\cdot 7075$, equals $2\cdot 1779 \times \cdot 324895$.

In view of the discrepancies between the first and second approximations it is necessary to carry out a second iteration, using in rows l_1 and l_2 the loadings obtained in rows l_1' and l_2' . The procedure is exactly as before. In the present instance it was desired to make quite sure of the presence or absence of a third factor; so a third iteration was carried out before testing the residuals for significance. The loadings from successive iterations are given later in Table VI on p. 00. It will be seen that the loadings have changed so little that the third iteration was not necessary.

The residuals are now tested for significance. Lawley's test involves the calculation of the quantity

$$N \sum \frac{\bar{r}_{ij}^2}{(1 - h_i^2)(1 - h_j^2)} = w \quad (\text{say}),$$

TABLE II. CALCULATION OF SECOND APPROXIMATIONS TO FACTORS I AND II (ITERATION 1)

Test No.	1	2	3	4	5	6	7	8	9	Check
BLOCK A	l_1	.7184	.7278	.5541	.7811	.7364	.7294	.4856	.8080	6.1923
	l_2	.2312	.1270	.2210	-.4360	-.4738	.2882	.1995	.3047	-.0121
	s^2	.4304	.4542	.6441	.1998	.3733	.3849	.7244	.2543	
BLOCK B	a_1	1.6691	1.6024	.8603	3.9094	1.7452	3.1578	1.8950	3.1773	
	b_1	10.0496	10.3281	7.5535	12.1488	10.1939	11.5206	10.1140	11.4934	90.0253
	c_1	9.3312	9.6003	6.9994	11.3677	9.5424	10.7842	9.3846	10.6854	83.8330
	l_1'	.6806	.7002	.5105	.8291	.6960	.7866	.6845	.4477	6.1146
BLOCK C	a_2	.5372	.2796	.3431	-2.1822	-1.2047	-2.0317	.7488	.2754	1.1982
	b_2	-.1732	-.5616	.0757	-3.0897	-2.8793	-3.2002	.1044	.0745	-.0206
	c_2	-1.3201	-1.3581	-.9902	-1.6081	-1.3500	-1.5257	-1.3277	-.8684	-1.5117
	d_2	.9157	.6695	.8449	-1.0456	-1.0796	-1.2007	1.1439	.7434	1.1864
	l_2'	.2975	.2175	.2745	-.3397	-.3508	-.3901	.3716	.2415	.3855

$k_1^2 = 187.9771$; $k_1 = 13.71047$; $1/k_1 = .072937$ $i_2 = -1.939594$ $k_2^2 = 9.47359$; $k_2 = 3.07792$; $1/k_2 = .324895$

TABLE V. CALCULATION OF SECOND APPROXIMATION TO FACTOR III (ITERATION 4)

Test No.	1	2	3	4	5	6	7	8	9	Check
BLOCK A	l_1'''	.6680	.6919	.4999	.8405	.7030	.8007	.4407	.7701	6.0848
	l_2'''	.3107	.2398	.2902	-.3158	-.3180	-.3679	.3912	.2481	.8976
	l_3	-.1313	.1698	.2392	.0472	.1789	-.1344	.0657	-.3497	+.0427
	s^2	.4400	.4349	.6087	.1916	.3727	.2055	.3937	.6219	
	a_3	-.2984	.3904	.3930	.2463	.4800	-.6540	.1669	-.5623	
BLOCK D	b_3	-.2027	.2580	.3724	.0768	.2760	-.1825	.0927	-.5823	.0423
	c_3	.0058	.0060	.0043	.0073	.0061	.0069	.0058	.0038	.0527
	d_3	-.0084	-.0065	-.0079	.0086	.0086	.0100	-.0106	-.0067	-.0243
	e_3	-.0688	.0887	.1368	.0137	.0824	-.0650	.0318	-.2297	-.0288
	l_3'	-.1193	.1539	.2373	.0238	.1429	-.1128	.0552	-.3985	-.0499

$i_3 = .0086694$ $j_3 = -.0271503$ $k_3^2 = .332307$; $k_3 = .576461$; $1/k_3 = 1.73472$

where N is the number of cases from which the original correlations are calculated, r_{ij} the residual of test i with test j , and h_i^2 the communality of test i ascribable to the factors already evaluated. The quantity w is approximately distributed as χ^2 with degrees of freedom $\frac{1}{2} \{ (n - m)^2 - n - m \}$, where n is the number of tests and m the number of factors extracted. In our example, $N = 211$, $n = 9$, and $m = 2$. The test of significance is satisfactory when N is large, and may be applied for about 200 cases or more. It is only valid when the factor loadings have been estimated by the present procedure. The calculation of w may be carried out systematically as follows.

The residuals are first computed in the usual way by subtracting the correlations due to the first two factors (as obtained from the third iteration in the present case) from the original correlations. It is not difficult to devise checks to ensure arithmetical accuracy of the work. These residuals, multiplied by 10^4 , are entered below the diagonal in Table IVA. The entries in the diagonal cells are the reciprocals of $(1 - h_i^2)$ taken from Table III. Above the diagonal of Table IVA the squares of the residuals, r_{ij}^2 , multiplied by 10^8 , are entered, and the sum of each row of these multiplied squares is entered in the column headed "Sum," the diagonal elements and the residuals themselves being of course excluded.

TABLE III. LOADINGS AND COMMUNALITIES FOR FIRST TWO FACTORS

Test No.	Factor Loadings		h_i^2	$\frac{1}{(1 - h_i^2)}$
	I	II		
1	.6680	.3107	.54276	2.1870
2	.6919	.2398	.53623	2.1562
3	.4999	.2902	.33412	1.5018
4	.8405	-.3158	.80617	5.1592
5	.7030	-.3180	.59533	2.4711
6	.8007	-.3679	.77647	4.4737
7	.6700	.3912	.60194	2.5122
8	.4407	.2481	.25577	1.3437
9	.7701	.4193	.76887	4.3266

TABLE IVA. RESIDUALS AND THEIR SQUARES

The residuals, multiplied by 10^4 , are given below the diagonal: their squares multiplied by 10^8 above the diagonal. The diagonal entries are values of $1/(1 - h_i^2)$ from Table III.

Test No.	1	2	3	4	5	6	7	8	9	Sum
1	(2.1870)	182	847	53	640	31	49	3881	29	5,712
2	-135	(2.1562)	4058	0	64	15	110	6642	123	11,012
3	-291	637	(1.5018)	724	121	1592	15	5446	8	7,906
4	73	2	269	(5.1592)	0	3	10	48	98	159
5	-253	80	110	-7	(2.4711)	0	1289	6740	1	8,030
6	56	-39	-399	17	-5	(4.4737)	416	2725	94	3,235
7	70	-105	39	31	359	-204	(2.5122)	59	0	59
8	623	-815	-738	-69	-821	522	-77	(1.3437)	708	708
9	-54	111	-29	-99	11	97	1	266	(4.3266)	---
	—	182	4,905	777	825	1,641	1,889	25,541	1,061	36,821

TABLE IVB. ROWS OF SQUARED RESIDUALS MULTIPLIED BY DIAGONAL ELEMENTS OF TABLE IVA

Test No.	2	3	4	5	6	7	8	9	Sum
1	398	1,852	116	1,400	68	107	8,488	63	12,492
2	—	8,750	0	138	32	237	14,321	265	23,743
3	—	—	1,087	182	2,391	23	8,179	12	11,874
4	—	—	—	0	15	52	248	506	821
5	—	—	—	—	0	3,185	16,655	2	19,842
6	—	—	—	—	—	1,861	12,191	421	14,473
7	—	—	—	—	—	—	148	0	148
8	—	—	—	—	—	—	—	951	951
Sum	398	10,602	1,203	1,720	2,506	5,465	60,230	2,220	84,344
	858	15,922	6,207	4,250	11,211	13,729	80,931	9,605	142,713

The last row is the product of the column sums of Table IVB and the corresponding diagonal elements of Table IVA.

The (squares $\times 10^6$) in each row of Table IVA are now multiplied by the diagonal element of the same row, and entered in Table IVB. The sums in the right-hand column of Table IVB should check against the product of the diagonal element and the corresponding sums in Table IVA. The columns of Table IVB are now summed, their grand total checking against the total of the last column. These column totals are then multiplied by the corresponding diagonal elements of Table IVA, and entered in the last row of Table IVB. For example, $858 = 2.1562 \times 398$. The sum of this last row, when multiplied by N and divided by 10^6 , gives w . We have $w = 211 \times 142,713 \times 10^{-6} = 30.112$. The degrees of freedom, $\frac{1}{2} \{ (n-m)^2 - n - m \}$, are 19, when $n = 9$ and $m = 2$. Entering the χ^2 table with d.f. = 19, we find the 5 per cent. significance level of χ^2 to be 30.144. The residuals are therefore just significant at the 5 per cent. level, and accordingly three factors are required to explain the original correlations. The same test of significance applied to the residuals after the second iteration gave $w = 30.356$, not appreciably different from the value previously found, thus showing that the third iteration was redundant.

These second residuals are now analysed by the centroid method, the highest correlation in each column being used as communality. Signs are changed as usual to give maximum variance to the third factor. Better estimates of this factor are thus obtained than from the residuals after removal of the first two *centroid* factors. The resulting loadings, l_3 , are then used as first approximations to the third factor. A second approximation to this factor is now obtained by iteration 4 as shown in Table V. The first two factors are not subjected to revaluation at this stage, since the specific variances of the tests are but little changed by the introduction of the third factor. Considerable time is thus saved.

The procedure is the same as for Factor II in Table II, except that one more stage is required.

1. In Block A, rows l_1''' and l_2''' contain the loadings in Factors I and II after the third iteration (the primes indicating the number of iterations), and row l_3 the loadings in Factor III given by the centroid analysis. Row s^2 is the specific variance, e.g., for test 1, $.4400 = 1 - (.6680)^2 - (.3107)^2 - (.1313)^2$.

2. Row a_3 is formed by dividing l_3 by the corresponding value of s^2 for each test. Thus for test 1, $-.2984 = -.1313/.4400$.

3. Row b_3 records the inner products of row a_3 with the successive columns or rows of the original correlation matrix, R . The calculation proceeds as for b_1 in Table II.

4. The quantity i_3 is obtained from the inner product of rows a_3 and l_1''' , and row c_3 is the product of row l_1''' and i_3 . This row corrects for the effect of Factor I and is later subtracted from row b_3 .

5. The additional stage now enters. The quantity j_3 is the inner product of rows a_3 and l_2''' , and row d_3 is the product of row l_2''' and j_3 . This row corrects for the effect of Factor II.

6. Row e_3 is (row b_3 - row i_3 - row c_3 - row d_3); thus for test 1, $-.0688 = -.2027 - (-.1313) - .0058 - (-.0084)$. In order to add (algebraically) three or more rows the use of a slotted paper as a guide is almost indispensable.

7. The quantity k_3^2 is the inner product of rows a_3 and e_3 . Row l_3' is row e_3 multiplied by $1/k_3$; it gives the second approximation to Factor III loadings.

The fifth iteration is now carried out. In Block A will be inserted the loadings l_1''' , l_2''' and l_3''' , row s^2 being the defect from unity of the sum of the squares of these three loadings. The working then follows the same course as Blocks B and C of Table II, and Block D of Table V. From Table VI it will be seen that the loadings of Factors I and II have not changed appreciably with this iteration, and that those of Factor III have settled down fairly well. A further two iterations have been carried out, however, to observe the progress of the loadings towards constancy.

If the second residuals had been highly significant, it would have been necessary to evaluate the third residuals and apply the significance test to them: in the present case it is obviously not necessary. We applied the test, however, after the seventh iteration and found $w = 211 \times .03773 = 7.961$. The degrees of freedom are 12. The value of χ^2 corresponding to $P = .50$ is 11.340; the third residuals are therefore far from significant. If fourth factor loadings had to be estimated, a centroid analysis of the third residuals would give a first approximation, and a second approximation would follow as done in Table V, Block A now containing four factor loadings for each test instead of three. In the process there would be four terms to subtract from row b_4 , the first three terms being formed analogously to those for Factor III in Table V; for the last term the inner product of a_4 and l_3' is formed, call it p_4 , and row l_3' is then multiplied by p_4 . A final iteration of all four factors together is then carried out.

We now show in Table VI the effect of successive iterations on the factor loadings in each test. (The analysis of the same correlations previously given by the author (1) was stopped at the fifth iteration: hence the slight deviations from the present loadings after seven iterations.)

TABLE VI. CONVERGENCE OF FACTOR LOADINGS

Test No.	Centroid Loading	Iteration No.						
		1	2	3	4	5	6	7
FACTOR I								
1	.7184	.6806	.6704	.6680	—	.6684	.6677	.6671
2	.7278	.7002	.6934	.6919	—	.6926	.6926	.6923
3	.5541	.5105	.5017	.4999	—	.5009	.4999	.4989
4	.7811	.8291	.8385	.8405	—	.8402	.8401	.8397
5	.6515	.6960	.7026	.7030	—	.7036	.7041	.7044
6	.7364	.7866	.7980	.8007	—	.8042	.8068	.8088
7	.7294	.6845	.6728	.6700	—	.6686	.6670	.6659
8	.4856	.4477	.4416	.4407	—	.4480	.4494	.4505
9	.8080	.7794	.7708	.7701	—	.7704	.7699	.7694
FACTOR II								
1	.2312	.2975	.3099	.3107	—	.3128	.3145	.3155
2	.1270	.2175	.2361	.2398	—	.2416	.2434	.2445
3	.2210	.2745	.2875	.2902	—	.2966	.2981	.2992
4	— .4360	— .3397	— .3190	— .3158	—	— .3139	— .3107	— .3079
5	— .4497	— .3508	— .3247	— .3180	—	— .3217	— .3216	— .3208
6	— .4738	— .3901	— .3708	— .3679	—	— .3717	— .3725	— .3733
7	.2882	.3716	.3886	.3912	—	.3897	.3899	.3904
8	.1995	.2415	.2476	.2481	—	.2658	.2708	.2741
9	.3047	.3855	.4118	.4193	—	.4211	.4229	.4241
FACTOR III								
1	— .1313	—	—	—	— .1193	— .1108	— .1052	— .1010
2	.1698	—	—	—	.1539	.1725	.1794	.1828
3	.2392	—	—	—	.2373	.2357	.2366	.2351
4	.0472	—	—	—	.0238	.0282	.0286	.0298
5	.1789	—	—	—	.1429	.1433	.1450	.1466
6	— .1344	—	—	—	— .1128	— .1073	— .1081	— .1079
7	.0657	—	—	—	.0552	.0579	.0587	.0603
8	— .3497	—	—	—	— .3985	— .4094	— .4249	— .4362
9	— .0427	—	—	—	— .0324	— .0193	— .0142	— .0112

III. COMMENTS ON THE METHOD

The present calculations were made with a Marchant calculator, fitted with automatic multiplication and division. One complete iteration for nine variables with three factors took $3\frac{1}{4}$ hours after experience of the method had been gained. For most analyses factor loadings correct to the second decimal place are all that are justified, and to judge from Table VI in a three factor analysis two iterations with two factors and two with three factors should be enough. This would occupy about 12 hours. To this has to be added the time spent on the preliminary centroid analysis, and that required for computing the second and third residuals and testing their significance. All things considered, 24 hours should be ample time. For more variables the time required is roughly proportional to the square of the number of variables. If more than three factors are evaluated, a greater allowance must be made.

It is well to estimate beforehand by one of the cruder significance tests how many factors are likely to be significant, and to put this number of factors, or one less, into the first iteration to avoid unnecessary tests of significance by Lawley's method. If the resulting residuals are significant, everything will have been gained. If not, then the iteration must be carried out with one factor less. If the loading of any test before and after iteration shows a large change, its loading for the next iteration may be judiciously altered in the direction of the change so as to hasten convergence.

The loadings given by Lawley's method frequently differ considerably from the centroid loadings. For example, in Test 6 of Table VI the first factor rises from $\cdot736$ to $\cdot809$, the second factor at the same time changing from $-.474$ to $-.373$. It is to be noted that a change in one direction of the first factor on iteration is almost always accompanied by a change in the opposite direction of the second factor, signs being disregarded as they are necessarily arbitrary. The loadings of the later factors may oscillate somewhat before settling down to constancy. It is occasionally advisable to estimate one factor more than is indicated by the significance of the residuals, for example when evidence in support of a previous enquiry is sought. The factors resulting from Lawley's method are orthogonal and may be rotated into new sets of factors, orthogonal or oblique according to taste.

IV. THE STANDARD ERROR OF FACTOR LOADINGS

With Dr. Lawley's kind consent and by courtesy of the Royal Society of Edinburgh we are able to give here Lawley's formula for the variance of a factor loading (6). The general formula for the variance of the r th factor in test i is

$$\sigma_{ir}^2 = \frac{\theta_r}{N} \left\{ 1 - \sum_{s=1}^r \theta_s l_{is}^2 + \frac{1}{2} \theta_r l_{is}^2 \right\},$$

where N is the number of cases in the sample, l_{is} is the loading in test i of the particular factor s , and $\theta_r = 1 + \frac{1}{k_r^2}$, k_r^2 being the quantity obtained in the computation of the factor loadings.

We now apply the formula to Factor III in Test 8, since it is obvious from Table IVB that this test contributes more than the other tests to the value of w . In this case the formula reduces to

$$\sigma_{8III}^2 = \frac{\theta_{III}}{N} \left\{ 1 - \theta_I l_{8I}^2 - \theta_{II} l_{8II}^2 - \frac{1}{2} \theta_{III} l_{8III}^2 \right\}.$$

From iteration 7 we have the following figures. $k_1^2 = 218.628$ (this figure corresponds to 187.977 in iteration 1); $k_2^2 = 10.6960$ (corresponding to 9.4736 in iteration 1); and $k_3^2 = .437,518$ (corresponding to .332,307 in iteration 4). Thus $\theta_I = 1.004,574$, $\theta_{II} = 1.093,493$, and $\frac{1}{2} \theta_{III} = 1.642,810$. From Table VI we have $l_{8I}^2 = .20295$, $l_{8II}^2 = .07513$, and $l_{8III}^2 = .19027$. Substituting in the above equation for σ_{8III}^2 , we find $\sigma_{8III} = .07906$, and $l_{8III}/\sigma_{8III} = .4362/.07906 = 5.517$. This quantity is approximately normally distributed, and the factor loading, l_{8III} , is thus highly

significant. It is important to note this result in relation to the fact that the second residual correlations were only on the verge of significance at the 5 per cent. level.

This test of significance only applies to the loadings before rotation and when the number of cases in the sample is large, say 200 or more. It ignores errors in estimating the communalities, but these errors are in general small, especially when the number of tests is moderately large.

V. THE STANDARD ERROR OF A RESIDUAL

Lawley (6) has derived a formula for the variance of an element of the residual matrix calculated from maximum likelihood factor loadings. The variance of the residual, $\bar{r}_{ij}$, is $\frac{1}{N} (e_{ii} e_{jj} + e_{ij}^2)$, where $e_{ii} = s_i^2 - \frac{l_{iI}^2}{k_I} - \frac{l_{iII}^2}{k_{II}} - \dots$, $e_{jj} = s_j^2 - \frac{l_{jI}^2}{k_I} - \frac{l_{jII}^2}{k_{II}} - \dots$, $e_{ij} = -\frac{l_{iI} l_{jI}}{k_I} - \frac{l_{iII} l_{jII}}{k_{II}} - \dots$, s_i^2 being the defect from unity of the communality of test i due to the factors already evaluated, l_{iI} the loading in test i of Factor I, and k the quantity evaluated in the course of analysis.

It has been difficult to maintain a uniform notation: in Table II an Arabic numeral was used to indicate the factor, while here it becomes necessary to use a Roman numeral. Thus l_2 and k_2 of Table II now become l_{II} and k_{II} respectively. Also e, i and j are used in a different sense from that in Tables II and V.

We can illustrate the use of the formula by testing the significance of the largest residual in Table IVA, $\bar{r}_{588} = -.0821$.

We have from Table VI $l_{5I} = .7030$, $l_{5II} = -.3180$, $l_{8I} = .4407$, $l_{8II} = .2481$, whence $s_5^2 = .4047$ and $s_8^2 = .7442$. Also $k_I = 13.982$ and $k_{II} = 3.0342$. These figures result from the third iteration and must be taken on trust. Substituting for these quantities in the equation for e_{55} , e_{88} and e_{58} , we find $e_{55} = .3361$, $e_{88} = .7100$ and $e_{58} = .0038$.

The variance of $\bar{r}_{588}$ thus becomes $\frac{1}{211} (.3361 \times .7100 + .0038^2) = .001131$ and its standard error, .03363. The residual, $\bar{r}_{588}$, is 2.44 times its standard error. This must be interpreted in relation to the total number of residuals in the matrix, 36 in all. Of these, on the basis of random fluctuations alone we should expect about two to exceed twice the standard error, disregarding sign. We might therefore hazard the opinion that $\bar{r}_{588}$ is only on the borderline of significance.

VI. SUMMARY

1. Lawley's maximum likelihood method of factor analysis has been illustrated by a practical example. Checks on arithmetical accuracy at each stage of working have been indicated.

2. Lawley's significance tests for a matrix of residuals, a single residual and a factor loading have been demonstrated.

REFERENCES

1. Emmett, W. G. (1949). 'Evidence of a space factor at 11 plus and earlier.' *Brit. J. Psych. (Stat. Section)*, II, p. 3.
2. Lawley, D. N. (1940). 'The estimation of factor loadings by the method of maximum likelihood.' *Proc. Roy. Soc. Edin.*, LX, p. 64.
3. Lawley, D. N. (1942). 'Further investigations in factor estimation.' *Proc. Roy. Soc. Edin.*, LXI, p. 176.
4. Lawley, D. N. (1943). 'The application of the maximum likelihood method to factor analysis.' *Brit. J. Psych.*, XXXIII, p. 172.
5. Slater, P., and Bennett, Elizabeth (1943). 'The development of spatial judgment and its relation to some educational problems.' *Occ. Psych.*, XVII, p. 139.
6. Lawley, D. N. (1949). 'Problems in factor analysis.' *Proc. Roy. Soc. Edin.*, Section A, LXII, p. 1.

ALTERNATIVE METHODS OF FACTOR ANALYSIS AND THEIR RELATIONS TO PEARSON'S METHOD OF 'PRINCIPAL AXES'

By CYRIL BURT

Department of Psychology, University College, London

I. *The Origin of Factorial Techniques.* II. *Highest Common Factors.* III. *Bipolar Analysis with Self-correlations of Unity: (a) Weighted Summation; (b) Simple Summation.* IV. *Bipolar Analysis with Reduced Self-correlations: (1) With Correction for Specificity; (2) Without Correction for Specificity: (a) Weighted Summation; (b) Simple Summation.* V. *Group Factor Analysis: (a) Non-overlapping Factors; (b) Overlapping Factors.* VI. *Summary and Conclusions.*

I. THE ORIGIN OF FACTORIAL TECHNIQUES

The Problem of Multiple Causation. There can be little question that the problem and fundamental principles of what psychologists call 'factor analysis' are due essentially to the pregnant suggestions of Galton and Karl Pearson. In his celebrated paper on 'Correlations and their Measurement' (1), Galton takes first the problem of concomitant variation between "two variable organs." After re-defining¹ the concept of 'correlation' as a measurable quantity, he points out that correlations may generally be regarded as "the consequence of the variations of the two organs being partly due to *common causes*." He explains in some detail how such correlations may be calculated, and then briefly considers the more general case in which a number of variables is involved. Where several unobservable causes are at work, he assumes that the resulting variation may be regarded as the *sum* of the "separate contributions"; and then discusses the possibility of estimating the "combined effect" of several observable variables, "whether these be themselves co-related or not." For this purpose he suggests 'transmuting' the observed measurements into terms of their probable error, adding the measurements so standardized, and "treating this set of values in exactly the same way as measures of a single variable" (1, p. 144). In this way we should evidently get what would nowadays be termed the 'general factor measurements' as calculated by simple or unweighted summation.

In several later papers² he deals with the treatment of marks or measurements obtained in examinations or other tests, and stresses the need for a more satisfactory

¹ In biology the idea of correlation is as old as Aristotle (e.g., *De Gen. Animal.*, II, 7; *De Part. Animal.*, I, 2-4, III, 14), who insists, in contradiction to the teaching of Plato, that classification must be based on the presence of a *number* of differentiae. 'Correlation' became the corner-stone of Cuvier's system of comparative anatomy; and in his view it was the aim of natural science, so far as possible, to convert 'empirical correlations' into 'rational correlations,' i.e., to refer observed covariations to some underlying function or cause of which each correlated group is the expression (*Hist. progr. nat. sci.*, 1826, I, p. 310; *Ossements fossiles*, 1812, I, p. 60f.). Galton seems to have taken both the term and the idea from Darwin (*Origin of Species*, II, Glossary, p. 310). His own contribution consisted (i) in recognizing that correlation need not be constant or perfect, and (ii) in proposing a quantitative procedure for dealing with such partial tendencies by measuring their relative strength.

² The most important of these is his paper on 'Assigning Marks for Bodily Efficiency in the Examination of Candidates for Public Services,' *Brit. Ass. Ann. Rep.*, 1889, pp. 471f. The problem there examined arose out of the investigations of a joint committee of the War Office and Civil Service Commission, appointed to enquire "whether the present literary examinations should be supplemented by physical competition." In support of his proposals, he quotes Edgeworth's paper (*J. Roy. Stat. Soc.*, LI, 1888, pp. 599-635; cf. *ibid.*, LIII, pp. 460f., 644f.) on 'The Unreliability of Judgements in the So-called Literary Type of Examination': (cf. also *Nineteenth Century*, 1889, pp. 303-8; *Proc. Roy. Soc.*, XLV, p. 145). For Pearson's comments on the suggestive but defective nature of Galton's proposals, see 18, II, pp. 334f., IIIA, p. 32f.

statistical procedure for combining such measurements. The relative weight of the component variables, he says, requires to be carefully determined, particularly when they are themselves correlated: but that is "a question of detail." Using an appropriate method of summation we may eventually arrive at "a measure of the degree in which one variable may be correlated with the *combined* effect of n variables." The result would be a much higher correlation, and a much better measure, than could be obtained by using a single variable.¹

Here he indicates the possibility of a mathematical procedure for deducing the correlation between a single trait and the hypothetical 'factor' common to a whole set of traits. Elsewhere (2, p. 132), in discussing the effect of inheritance on the resemblance between different members of a family, he considers how to deduce the expected correlation between two traits (a and b say), when we are given the hypothetical correlation of each with the common factor or cause (x say). He concludes² that the correlation (or regression) can be predicted by taking the product of the two hypothetical correlations with the factor, i.e.,

$$r_{ab} = r_{ax}r_{bx} \quad (i)$$

Recalling popular phrases about 'pure and mixed blood,' he compares the combination of different family strains to the mixture of different fluids of varying degrees of saturation (a metaphor which has provided a useful technical term). "The effect," he says, "resembles that of pouring a measure of water into a vessel of wine: the wine is diluted to a constant fraction of its original strength, whatever that strength may have been." Thus, when we seek to measure the diminishing degrees of resemblance between members of a family that are more and more remote, the result is "to weaken the original strength in a constant ratio," and "the process throughout is one of *proportionate dilutions*" (2, pp. 105-6). Pearson, however, points out (16, p. 38) that the equation for r_{ab} , in the foregoing simple form, would imply that any partial correlation due to other causal factors necessarily

vanishes (e.g., that $r_{ab.x} = \frac{r_{ab} - r_{ax}r_{bx}}{\sqrt{1-r_{ax}^2}\sqrt{1-r_{bx}^2}} = 0$). Hence Galton's product formula requires to be generalized; and this, he says, demands a more rigorous investigation of the problem of multivariate correlation.³

As is well known, in dealing with bivariate correlation Galton noted that "all sections of the surface of frequency," formed by two correlated variables, would constitute "a series of similar and concentric ellipses." His own geometrical approach is illustrated by his diagram of a correlation ellipse (2, p. 101, Fig. 11, reproduced by Pearson, 16, p. 36). He calculates the lengths and inclination of the principal axes; and, assuming that their lengths are proportional to the common cause and the cause of error respectively, shows how they are related to the resulting correlations, regressions, and errors of arrays: (*ibid.*, p. 103f.; cf. p. 223f.). Pearson generalizes these results to the case of

¹ Stern seems independently to have developed much the same notion. In our endeavours to measure *Grundeigenschaften*, he suggests, we should "aus den Einzelsymptomen auf synthetischem Wege ein neues Symptom schaffen, das jene beträchtlich an Symptomwert übertrifft"; and he cites in a post-script my own 1909 investigation (11) as one which "schon eine erstmalige Anwendung dieser Methode bringt" (*Differentielle Psychologie*, 1911, p. 294).

² His illustrative table (2, p. 133) gives what is virtually a 'hierarchy' of coefficients (each calculated by the above 'product formula') to represent the 'hierarchy of kinship' which results when we assume that a single general factor, viz., a common heredity, is operative throughout. It will be seen that his figures clearly exhibit the law of 'proportionate dilutions,' particularly if the reader continues the sequences, as Galton suggests.

³ The generalized form of the product theorem, $r_{ab} = r_{ax}r_{bx} + r_{ay}r_{by} + \dots$ (where $x, y, \dots$ denote uncorrelated factors), was not, I think, given by Pearson himself; but it is readily deduced from Pearson's geometrical treatment of partial correlation, and from the analogies that he elsewhere points out with the parallelogram law in dynamics. It implies that we must examine successive sets of residual correlations ($r_{ab} - r_{ax}r_{ay}$, etc.) for evidence of other possible factors (such as y); and thus furnished the basis for early attempts at multiple factor analysis. The practical use made of the procedure was sharply criticized at the time. But, under the title of the 'cosine law,' the generalized theorem was formally proved by Pearson's assistant, Maxwell Garnett (15, 1919, p. 96, eq. 16). Thurstone has since named it the 'fundamental factor theorem' (*Vectors of Mind*, 1935, eq. 1, p. 92).

multiple correlation (4); and, as we shall see, both he and Edgeworth proposed to treat the axes of the multiple 'correlation ellipsoid' as representing hypothetical 'index characters' or factors (5, 8, 9).¹

By 1889, as Pearson observes (15), "Galton had completed the theory of bivariate normal correlation." "The next stage in the theory" was the development of the principles of "multivariate correlation": "Galton endeavoured to reach this by a short cut." But here he was gravely handicapped, because "the geometry of n dimensions, which the theory of multiple correlation involves, remained for him a closed book" (18, II, pp. 380f.). Until this further problem had been solved, any rigorous method of factorial analysis was out of the question. By 1897 Pearson tells us, the solution had been worked out by himself (4) and Sheppard.² Shortly afterwards Yule, at that time Pearson's assistant, succeeded in dispensing with the assumption of normality by deducing the same formulæ from the principle of least squares (6, 10). This, so Pearson maintains, involved no fundamental change, since minimizing the sum of the squares of the residuals is virtually the same as 'maximizing the generalized correlation coefficient' (16, p. 37f.).

Now if, as Galton suggests, the best mark for a man's 'efficiency' is to be obtained by weighting standardized examination marks, the question at once arises: how are we to determine the weights? When external evidence is available (derived, for example, by following up the subsequent progress of the candidates), that might provide a possible criterion. But with most examinations the only accessible information is the internal evidence furnished by the results of the examination itself.

In its initial form the primary object of the multivariate procedure was to secure the best multiple correlation with some 'external criterion.' Pearson, however, also observed that the same essential principles could be applied to reach a satisfactory estimate when nothing but 'internal' evidence was procurable. Once again the method of least squares was invoked; and the outcome, as we shall see in a moment, was the proposal to choose the weights in such a way that the new composite variable would yield the best possible fit to the marks or measurements actually observed. The sums of the standardized measurements, appropriately weighted in this way, provide an improved estimate for what we should now call the first or 'general' factor—an estimate obviously more satisfactory than Galton's original 'short cut.' The multivariate procedure further enables us to estimate the supplementary factors, which Galton had treated as negligible. Throughout, the basic assumption is that "the unobservable variables" (i.e., the 'factors' which we are seeking to estimate) "may be supposed to be *uncorrelated causes*, connected by unknown (but determinable) functional relations with the correlated variables" (16).

In various memoranda and articles I have briefly described Pearson's final³ method of reducing correlated measurements to terms of a hypothetical set of uncorrelated components; and I have argued that practically all the methods of factor analysis used in this country are really simplifications or modifications of this funda-

¹ At the time they were writing, Czuber's well-known *Theorie der Beobachtungsfehler* (Leipzig, 1891)—a treatise invaluable to all who were interested in statistical analysis at the turn of the century—had recently appeared. Following Schols and others, he gives an admirable discussion of the possibility of resolving such 'errors' into orthogonal (*zueinander normale*) *Componenten*, on the lines of the parallelogram of forces (Fig. 1, p. 372), and gives working procedures for calculating the *Hauptachsen der Fehlerellipsoid*, on the analogy of the calculations used in determining the *Hauptachsen der Trägheit*. I imagine both Edgeworth and Pearson, who often quote this book in other connections, largely derived their concepts and their methods from this suggestive discussion.

² Sheppard, W. F., *Phil. Trans.*, CXII, 1898, pp. 101–167: cf. *id.*, 17. A simpler version of Pearson's proof, "based on a lecture-demonstration by Prof. Pearson," is given by Brown (12, pp. 128–130: Pearson's equation for the "ellipsoid in n -dimensional space" there appears as equation iv; the *m* "absolutely independent causes" may be regarded as factors). It seems a pity that this chapter was omitted in later editions, as the original form of Pearson's proof is difficult for the psychological student to follow.

³ Pearson seems to have considered other possible procedures. Thus, in discussing Galton's proposal to sum standardized marks or ranks, he observes: "ranking in multiple marking can be thrown back, through multiple correlation, on the P and χ^2 of the 'Goodness of Fit' tables." The article on the 'History of Correlation' (from which I have quoted in the text) refers primarily to Galton's work on heredity; but it seems equally applicable to Galton's other problems. I may add that the section from which the quotation is drawn gives a sufficient reply to those who have argued that Pearson's own contribution was nothing but a revival of the proofs of Gauss and Bravais.

mental procedure (e.g., 23, pp. 289–294 and refs.). To-day, however, as a reference to books on factor analysis will show, the contributions of Pearson and his fellow-workers seem either to be silently taken for granted or else to be entirely overlooked.¹ Nor, so far as I am aware, is any clear explanation available, showing the relations and the differences between his procedure and the later attempts that grew out of it. Accordingly, in the present article I propose first to describe at greater length the technique that he himself put forward, and to illustrate it by an actual application to the problem that he himself had in mind ; then to compare the results so obtained with those that would be furnished by more recent procedures ; and finally to discuss the reasons for these later modifications, and the theoretical and practical difficulties that they were intended to meet.

The Problem of Multiple Classification. Before describing Pearson's formulæ and proofs, a word or two is necessary about the context of thought within which his suggestions were developed. His paramount interest was in the investigation of what he called 'the factors of evolution.' This required the study and measurement not of individuals, but of types.² The multiplicity of types, so he held, rendered it impossible to account for the observable facts by means of a single 'inherent growth-force,' as certain current theories assumed : their variety suggests a far more complex scheme of causation. Thus the problems of multiple classification and of multiple

¹ In the early discussions of factorial procedures in psychology during the years 1905 to 1915, the statistical proposals of Galton and Pearson were constantly cited, sometimes as suggesting ideas for imitation or development, sometimes as furnishing matter for criticism and replacement. At the symposium on mental testing, arranged at the Sheffield meeting of the British Association in 1910, I defended my tentative methods on the ground that they were an endeavour to apply to psychological problems "the statistical devices put forward so suggestively by the biometric school for dealing with physical characters" ; the method of summation was advocated, as a simplification of Pearson's procedure, for calculating the 'highest common factor' ; and Galton's authority was cited for the product-theorem (used in reconstructing the best fitting hierarchy ; cf. 11, p. 160f.). Dr. Brown criticized my modifications on the ground that Pearson's full method of partial correlation should have been employed (cf. 12, p. 93) ; Professor Spearman claimed that his correction formulæ were based on the equation for partial correlation (Yule's rather than Pearson's), but insisted that the pattern of correlations obtained with physical measurements was so different from that obtained with mental tests, that, in the psychological field, at any rate, it was no longer possible to "take the methods of the Galton-Pearson school as a model" (cf. *Am. J. Psych.*, XV, p. 97). During the subsequent discussions of the British Association Committee on 'Mental and Physical Factors,' and at later meetings of the British Psychological Society, a sharp opposition emerged between the views of Pearson on the one hand and Spearman on the other. Nevertheless, in a few short years all except the main protagonists appeared to have accepted an intermediate position between the two extremes : (cf. 19, pp. 225f.).

Writers, who entered the field at a later date, were naturally unfamiliar with these early developments, particularly as newer controversies largely replaced the old. Thus neither of Thurstone's books (*Vectors of Mind* and *Multiple Factor Analysis*) contains any reference to Pearson. Thomson's *Factorial Analysis of Human Ability* refers only to Pearson's work on 'univariate' and 'multivariate selection' and on sampling errors in the 'theory of a generalized factor.' Holzinger and Harman merely cite the paper of Pearson and Filon on the correlation between the errors of correlations. Thurstone, it is true, now seems to favour "the principal axes solution," and believes that, when the difficulties of computation are overcome, "it will be preferred by all students of this subject." But he states that this solution was first suggested to him in 1931 by Professor Bartky, and seems quite unaware of its earlier advocates (*Multiple Factor Analysis*, pp. 473–4, 509). In this country most psychologists now credit the idea of 'principal axes' to Thurstone or (more commonly) Hotelling (cf. 27, p. 68), and appear unaware of the earlier uses and references by British writers.

² "If we took the modes for a great variety of characters, we should have a *type*" (7, p. 382). The 'type,' like the 'class,' is thus based on a *number* of associated traits ; but, unlike the 'class,' it does not involve discontinuity or sharply demarcated boundaries. (Cf. Stern, *Differentielle Psychologie*, p. 173. Stern very appropriately quotes Sigwart's definition of 'type' and his arguments for the necessity of some such concept for purposes like the classification of men according to their body-build : *Logic*, II, pp. 520–1).

causation prove to be closely connected. But to Pearson, as to Galton, the problem of classification seemed to come first.¹

It may be recalled that the idea of measuring correlations was originally introduced by Galton, not (as is so often stated) as a method of investigating inheritance, but as an aid to classifying individuals on the basis of standardized measurements.² By the operation of heredity and other causes, he believed, certain 'stable forms' were developed, the members of which present a complex pattern of resemblances, due to clusters of correlated traits. The extent of the resemblances might be either broad or narrow. As a result, he declares, "Personal Forms . . . may be divided into Types, Sub-types, and Deviations from them"—a threefold division which re-emerges under various names again and again in the subsequent literature.³

The need for a satisfactory method of classifying persons according to a scheme of physical types cropped up as a result of a practical demand—namely, the desire to identify criminals by means of a brief set of body-measurements, such as would serve to assign each individual to a place in a convenient anthropometric filing-system. The Home Office had provisionally adopted the procedure suggested by Bertillon.⁴ Galton, however, maintained that several of the twelve measurements prescribed in Bertillon's schedule were so closely related as virtually to duplicate each other, while those that were relatively independent were too few to furnish a sound classification. Plainly, if we have measured Arsène Lupin's left leg, we shall not add much to our information by measuring his right leg as well. But Bertillon still insisted on measuring one of his arms. Galton held that arm-length would vary with leg-length almost as closely as the length of one leg varied with the length of the other, and proposed to calculate the degree of resemblance in order to substantiate his contention. From his figures it seemed clear that anthropometry must distinguish between 'unlike organs' (such as skull and collar-bone, or head-length and height) for which the correlations will be relatively low (about 0.35 according to Galton), and 'like organs' (such as legs, arms, and fingers) for which they will be decidedly high (0.70 to 0.85). He concludes that Bertillon's selection of traits is neither adequate nor economical. What is needed is a set of representative traits, so chosen that each trait shall have the *smallest possible correlation* with each of the others: in this way a reliable set of index-characters will be obtained with the fewest number of traits.

Classification by Means of Uncorrelated Characters. A few years later in a brief but instructive *Note*, Edgeworth⁵ took up the problem, and put Galton's conclusions and suggestions in a more rigorous and explicit form. He begins by describing the actual method of classification adopted by "the anthropometer who occupies the Bureau of Identification at Scotland Yard." For every trait the whole range of variation is divided into three sections with an equal number of persons in each: i.e., the measurements are grouped under three headings—"small," "medium," and "large." In theory, with n independent traits this should yield 3^n compartments.

¹ Cf. 7, pp. 377f. Pearson, however, laid considerable emphasis on the study of change. And I am tempted to suggest that this aspect has been unduly ignored by psychologists in subsequent factorial work. The reason no doubt is that a static analysis must almost inevitably precede a dynamic.

² This is stated more than once by Galton himself: cf. 'Human Variety' *Pres. Add., J. Anthropol. Inst.*, XVIII, 1889, pp. 401-19; *Anthropol. Lab Notes*, 1890, p. 10f.

³ 2, pp. 25f. Cf. *Brit. Ass. Ann. Rep.* (1885), pp. 1206f. The idea of classifying persons according to a scheme of (i) 'primary types' (genera), (ii) 'subordinate types' (species), and (iii) 'individual deviations from them', is in effect an early statement of the three factor theory: (cf. 19, p. 19).

⁴ Bertillon, A., *Signaletic Instructions, including the Theory and Practice of Anthropometrical Identification* (Eng. trans., 1896). The 'metric system of identification' in force at that time at Scotland Yard is fully described by Dr. Garson, *J. Anthropol. Inst.*, XXX, 1900; and, as Pearson points out, both Bertillon and Garson "claimed that the measurements for the different characters would be independent" (cf. 18, IIIA, p. 5). Galton later urged the substitution of finger-prints for 'bertillonage'; and his interest reverted almost exclusively to the more theoretical study of human inheritance.

⁵ 5, pp. 534-7, 539. (The 'leg' measurement was 'height of knee above ground.') Dr. Sophie Bryant (Headmistress of the Camden High School, London, the first, I believe, to apply Galton's method of testing to school children) gives a simplified procedure for calculating Edgeworth's 'correlated averages' (3). But, disappointingly enough, she applies it not to her own test-results, but to "measurements of the organs of shrimps."

There is no need here to decide claims to priority. In an earlier paragraph of the *Note*, Edgeworth refers to "Pearson's splendid series of Mathematical Contributions to the Theory of Evolution," and to the fact that Pearson was the first to point out "the most probable determination" of the constant r , which figures in Edgeworth's formula, and which Edgeworth admits he arrived at by a method "of a somewhat arbitrary character."

But in practice, owing to the high intercorrelations of many of the traits, "the anthropometer finds himself, with respect to several of his compartments, in the position of Old Mother Hubbard." To overcome this defect Edgeworth proposes to carry Galton's principle to its furthest limit, and to substitute a set of hypothetical characteristics which have *no intercorrelation* whatsoever.

He deals first with the bivariate case ; and prints (in an appendix) a simplified formula for finding the major and minor axes of the correlation-ellipse (very similar to that already used by Galton). He then states : " This principle may be extended to any number of attributes : there are always as many *independent characteristics* as attributes." These are to be found as follows. " The exponent of the [multivariate] frequency function is a quartic of which the coefficients involve r_{12} , r_{13} , r_{23} , etc. . . . ; and this quartic is to be transformed to *principal axes* according to well-known rules." To illustrate the suggestion he gives weighting equations, 'roughly worked out,' for three 'characteristics' (or 'factors,' as we should call them), derived from Galton's correlations for three absolute body-measurements (in inches), viz. :

$$\text{No. I} = 0.16 \text{ stature} + 0.51 \text{ forearm} + 0.39 \text{ leg} ;$$

$$\text{No. II} = -0.17 \text{ stature} + 0.69 \text{ forearm} - 0.09 \text{ leg} ;$$

$$\text{No. III} = -0.15 \text{ stature} - 0.25 \text{ forearm} + 0.52 \text{ leg}.$$

We see that the 'characteristics' or factors are weighted sums. 'Characteristic No. I' is evidently a 'factor for general body size' ; No. II will pick out those who are disproportionately long or short in the arm ; No. III those who are disproportionately long or short in the leg.¹

Meanwhile, from 1895 onwards, in his numerous 'Mathematical Contributions to the Theory of Evolution,' Prof. Karl Pearson had been working systematically on the various biometric problems raised by Galton ; and his results influenced the early development of statistical psychology more than those of any other writer. One of his first collaborators, Dr. W. R. Macdonell, secured from the Central Metric Office, New Scotland Yard, figures for seven physical traits measured on 3,000 criminals in accordance with Bertillon's methods. The data were analysed in the Galton Laboratory, and the results examined in a long and important paper (9). Here for the first time we find a complete table of coefficients, showing the correlations of every variable with every other, printed in the now familiar form of a symmetrical determinant or matrix (*loc. cit.*, pp. 202, 207).

In Table I below I have reproduced the figures to four decimal places only. In the original table the figures in the leading diagonal are given as 1.000 throughout : for reasons which will appear in a moment I have substituted the 'reduced self-correlations,' which are placed in brackets. I propose to use this table² to illustrate

¹ Edgeworth's account of his procedure is highly condensed. I have set out what I take to be his method of calculation in *Notes on Factor Analysis* (1930 : reworked) ; and am indebted to Mr. L. T. Simpson (who attended his lectures) for the explanation there given. Here as elsewhere Edgeworth started from the principle that we should seek those hypothetical constants, which would maximize the 'probability of the given observations' (i.e., their 'likelihood'), instead of seeking (as he says Gauss and Laplace preferred to do) those estimates which would 'minimize the squares of the errors' (cf. 3, p. 190, and *J. Roy. Stat. Soc.*, LXXI, p. 385) : in other words, he prefers Gauss's *Theoria Motus* approach, i.e., Gauss's first derivation of the 'best measurement or linear combination' (*loc. cit.*, II, iii, §177), to his later *Theoria Combinationis* approach (1821, art. 38), i.e., his 'second proof' : (see Czuber, pp. 237, 289, esp. letter to Bessel of 1839). Edgeworth later stated the conditions under which the two solutions would coincide (*loc. cit.*, p. 386). For the present problem he proposed to assume that "the moduli for the fluctuations of each attribute, and therefore the precisions, should be taken as unity."

² I have, as a matter of fact, regularly used the table in this way in my *Notes on Factor Analysis*, and recommend it for that purpose. It has many advantages. First, it is easier for the beginner to visualize and think about physical traits and their relations than about mental traits. Secondly, physical measurements are subject to much smaller errors of measurement ; and the analysis of correlations based on several thousands of cases is relatively unhampered by the problem of sampling error. Thirdly, the table shows several peculiarities which raise in a clear-cut way the possibility or desirability of different modes of analysis. And finally this was the first table which it was suggested might be analysed into a series of uncorrelated components or 'factors' ; it marks, as it were, the birth of factor analysis as a biometric method.

the results obtained first by Pearson's own procedure and then by the various methods of factor analysis that have developed out of his proposals.

TABLE I. CORRELATIONS BETWEEN PHYSICAL MEASUREMENTS (Macdonell)

Trait	Head L.	Head B.	Face	Foot	F'arm	Height	Finger
1. Head Length ..	(.2998)	.4016	.3945	.3389	.3054	.3399	.3007
2. Head Breadth ..	.4016	(.7675)	.6178	.2061	.1352	.1831	.1504
3. Face Breadth ..	.3945	.6178	(.5688)	.3632	.2887	.3453	.3210
4. Foot	.3389	.2061	.3632	(.7440)	.7970	.7364	.7587
5. Forearm	.3054	.1352	.2887	.7970	(.9990)	.7999	.8464
6. Height	.3399	.1831	.3453	.7364	.7999	(.8595)	.6608
7. Finger	.3007	.1504	.3210	.7587	.8464	.6608	(.8498)
<i>First Factor-Saturations</i>							
Macdonell's Method	.4861	.4867	.5050	.7265	.9013	.6583	.7326
Pearson's Method ..	.5382	.4130	.5751	.8777	.8884	.8493	.8539
Burt's Method ..	.4435	.3792	.5101	.8453	.9229	.8452	.8496

II. HIGHEST COMMON FACTORS

The Method of Principal Axes. After a full consideration of certain preliminary questions,¹ Macdonell turns to the problem of classification; and quotes at some length the suggestions put forward by Pearson for calculating "ideal index-characters." Here, and in a supplementary paper published in his own name, Pearson expounds a method of transformation (very similar to that of Edgeworth) which formed the starting-point for various simplified suggestions put forward soon afterwards by those who were concerned with the problem of psychological classification.

Like Edgeworth, Pearson begins by assuming that what we really need are what we should now call uncorrelated or 'orthogonal factors.' However, unless some further restriction is imposed, it is evident that there are innumerable ways in which n correlated variables can be expressed in terms of n uncorrelated components. The solution can be made determinate by introducing a further condition, which has intrinsic advantages of its own. In practice, the best order for drawing up the classificatory set would plainly be to take each item in the order in which it helps to reduce individual variability: in the interests of economy, therefore, it is suggested that the first of the hypothetical index-characters should be so determined that it will reduce variability as much as is possible at a single step, and similarly with each of the remaining characters in turn.

In modern terminology, this means that we must select for our first and most important index-character that particular component which will account for the maximum amount of variance;

¹ At that date the chief obstacle to the general introduction of the correlational technique was the seemingly laborious calculations required by the full product-moment method. The first part of Macdonell's paper is largely concerned with demonstrating the practical accuracy of a 'new short method' (Pearson's tetrachoric correlation). In subsequent references to the paper, this is the point commonly noted; and the suggestion of factorial analysis was consequently overlooked.

In considering the fit obtained by various methods of factorization, it is to be remembered that we are dealing with tetrachoric correlations, some of which have apparently been corrected by the fuller procedure. Hence the standard error of an ordinary product-moment coefficient, based on a sample of 3,000, will be inapplicable. Macdonell suggests that the correlations can be regarded as equivalent to product-moment figures obtained from 1,100 individuals.

and for the second index-character that particular component which will account for a maximum of the residual variance; and so on. It follows that the last index-character will be that which has the least possible variance, i.e., that for which the sum of the squares of the factor-measurements is smallest. Hypothetical components, calculated according to this principle, I ventured to call 'highest common factors' (cf. 8, 1909): each is made to yield in its turn the highest set of correlations with the several traits that have been measured.

A system of index-characters, conforming with these requirements, would, so Pearson contends, furnish the best descriptive scheme of all. He explains the calculation required by pointing out that these "ideal index-characters would be found, if we calculated for the directions of the seven uncorrelated variables *the principal axes of the correlation ellipsoid*. Thus, given $x_1, x_2, \dots, x_7$ as the correlated variables, the seven uncorrelated variables would be:

$$\left. \begin{aligned} X_1 &= l_{11}x_1 + l_{12}x_2 + \dots + l_{17}x_7, \\ X_2 &= l_{21}x_1 + l_{22}x_2 + \dots + l_{27}x_7, \\ \text{etc.,} \end{aligned} \right\} \quad (\text{ii})$$

where the l 's give the direction cosines of the *principal axes*." "Once the preliminary determination of the 49 numerical multipliers had been calculated," he adds, "the uncorrelated characters of a criminal could easily be found from the measured characters."¹

This is the first explicit outline of a formal procedure for carrying out a factorial analysis. It will be noted that the 'index-characters' or 'factors' are defined by the above equations as *weighted sums of the trait-measurements*, and that the principle indicated for calculating the requisite weights implies that the 'factors' so envisaged are (a) multiple, (b) orthogonal, and (c) determined so as to fit the data in accordance with the principle of least squares. In fact the method proposed is identical with what has since become known as 'factor-analysis by least squares' or 'weighted summation.'

The Lines of Closest Fit. In a highly suggestive paper (8), published a little earlier, Pearson had already set out in a fuller and more general form the mathematical basis for the procedure.² In developing his arguments he adopts a practice that has become almost universal in factorial problems, namely, the assumption that the relations between the data may be formulated in terms of the 'geometry of hyperspace,'³ following (so he implies) the conventions adopted in describing the composition and resolution of forces in the theory of dynamics.

He begins by pointing out that "in many physical, biological, and statistical investigations it is desirable to represent a system of points, in two, three, or higher dimensional space, by means of the 'best-fitting' line or plane." "Best fit" he defines as that which will "make the sum of the

¹ *Loc. cit.*, p. 209. In 'higher algebra' the transformation effected by a system of non-homogeneous equations such as (ii), and obeying the conditions specified by Pearson, is commonly known as a 'principal axis transformation': cf., for example, MacDuffee, C. C., *Vectors and Matrices*, 1943, pp. 171. Adopting the matrix notation suggested in 24, p. 71f., Pearson's equations can be briefly written $X = L'M$, where $X = \Lambda^{\frac{1}{2}}P$ is the matrix of unstandardized or non-normalized factor-measurements (cf. also 23, p. 487f.: to avoid confusion I use M to designate the matrix of trait-measurements which Pearson denotes by the variables $x_1, x_2, \dots, x_7$).

² R. J. Adcock (*Analyst*, V, 1878, pp. 53-4) and M. Merriman (*Method of Least Squares*, 1885, p. 127) had briefly dealt with the simpler bivariate problem, arising "when *both* observed quantities are affected by error." Pearson's treatment is fuller and more general than that of any of his predecessors. His paper has thus become the *locus classicus*, and has given rise to numerous discussions since: for refs., see Roos (25). Roos and others point out that Pearson's result is not invariant under homogeneous strain (e.g., change of scale), and that it can have no special value unless the data are first put into terms of their standard deviations. This of course was expressly done in applying the method to the determination of 'factors.'

³ In the *Grammar of Science*, he observes: "That a quantitatively exact study of defective children should need the study first of the geometry of hyperspace may sound paradoxical; but it is none the less true" (7, p. 411).

squares of the perpendiculars from the system of points upon the line (or plane) a minimum.”¹ Using the same notation as before, this implies that we must minimize the sum of the squares of the deviations, which he writes $\sum (l_1x_1 + l_2x_2 + \dots + l_qx_q - p)^2$, subject of course to the condition that the estimated points shall lie on the same straight line (or plane). He follows Lagrange’s procedure ; introduces an undetermined multiplier, Q ; and then shows that in order to reach a non-trivial solution for the direction-cosines ($l_1, l_2, \dots, l_q$) we have to solve a determinantal equation which he prints as follows :

$$\begin{vmatrix} 1 - \frac{\sum^2}{\sigma_1^2} & r_{12} & \dots & r_{1q} \\ r_{21} & 1 - \frac{\sum^2}{\sigma_2^2} & \dots & r_{2q} \\ \dots & \dots & \dots & \dots \\ r_{q1} & r_{q2} & \dots & 1 - \frac{\sum^2}{\sigma_q^2} \end{vmatrix} = 0 \quad (\text{iii})$$

where $\sum^2 = Q/n$. He thus demonstrates “ that the line which best fits a system of points in q -fold space will pass through the *centroid* of the system, and will coincide in direction with the least *principal axis* of the ellipsoid of residuals ” or, what amounts to the same thing, with the largest *principal axis* of the “ correlation-type ellipsoid.”²

If we require supplementary lines to describe the system (i.e., supplementary factors to fit the residuals remaining after the first or subsequent factors have been partialled out), then the other ‘ principal axes ’ of the ellipsoid will afford the best representations of the remaining ‘ dimensions.’ All these lines are at right angles to one another. Hence “ physically the axes of the correlation-type ellipsoid are directions of *independent or uncorrelated variation*.” Thus, we have at length reached an exact expression for reducing the q observed and correlated variables to terms of q hypothetical and uncorrelated ‘ index-characters.’

In my *Notes on Factor Analysis* (1930) I endeavoured to simplify the foregoing proof by using matrix notation. Let the n measurements obtained by any given person be denoted by the vector m (n = Pearson’s q) : then m may be treated as giving the co-ordinates of a point in an n -dimensional space. And let the general factor we are trying to measure be represented by a line in that space whose direction-cosines are l ; then the square of the distance from the point to the line will be $m'm - l'mm'l$. Since mm' is fixed, minimizing this square will be equivalent to maximizing $l'mm'l$; and minimizing the sum of all such squares will mean maximizing $l'MM'l$. But, if the

¹ The definition is avowedly arbitrary. But it can readily be justified along the lines indicated by Edgeworth and later adopted by Miss Dent (22, pp. 23f.). “ To avoid the bugbear of inverse probability,” she starts by expressing the compound ‘ probability ’ of the errors in terms of ‘ Gaussian error functions,’ and observes that this expression may be regarded as “ a function of the unknown quantities ” (i.e., as a ‘ likelihood function,’ though this phrase is not actually used, doubtless because she is writing primarily for physicists). She then argues that “ by making this a maximum, subject to the condition that the points are collinear, we shall obtain the parameters of the required line.” Like Edgeworth, she uses the Gaussian ‘ modulus of precision ’ (h), rather than the standard deviation ; and eventually obtains a formula for the inclination of the required line, which may be

written $\tan 2\phi = \frac{2h_1h_2\sigma_1\sigma_2r_{12}}{h_1\sigma_1^2 - h_2\sigma_2^2}$. In the ‘ special case ’ in which the two precisions are equal, this reduces to Pearson’s equation for the inclination of the axis : (cf. Yule and Kendall, p. 231, eq. 12.10). She then takes up the problem of “ estimating the value of the precision constants from the given data,” when these consist of only single (i.e., unrepeatd) observations ; and deduces formulae for the errors in estimating the centroid and the inclination. Miss Dent is concerned solely with the bivariate case ; but her essential arguments can easily be generalized.

² The student often inquires how the ‘ line of closest fit,’ as here defined, differs from the more familiar ‘ regression lines,’ which are also frequently described as ‘ lines of closest fit.’ The answer is that (as Pearson makes clear) the regression line assumes that only the dependent variable is subject to error, whereas we are now concerned with the case in which *both* variables are subject to error. Hence, in determining a regression line, the distances whose square-sum is to be measured are distances measured in a direction parallel with one or other of the axes of co-ordinates ; in determining the principal axis they are measured in a direction orthogonal to (i.e., perpendicular to) the line itself, since this direction gives the shortest distance from each point to the line. The distinction is brought out clearly by Cramér, who proposes to call the line so determined the ‘ orthogonal regression line ’

measurements are in unitary standard measure, $MM' = R$, the matrix of correlations with unity in the diagonal. Hence our problem is ¹ to choose l so as to maximize $l'Rl$, subject to the condition that $l'l = 1$. Introducing a Lagrange multiplier λ (Pearson's Q), we form the expression

$$l'Rl + \lambda l'l, \quad (\text{iv})$$

Partially differentiating this with respect to each of the direction cosines in turn, and setting the derivatives equal to zero, we obtain

$$(R - \lambda I)l = 0, \quad (\text{v})$$

a set of n linear equations for finding the direction cosines. These equations will have a non-trivial solution only if the characteristic function

$$|R - \lambda I| = 0. \quad (\text{vi})$$

We thus have first to solve vi (Pearson's equation, iii above, in matrix notation) to find λ ; and then, taking each of the n roots in turn, to solve v for the corresponding values of l . It is easy to show that the row vectors $x_1, x_2, \dots, x_n$ thus obtained will be mutually orthogonal: so that $\Lambda^{-\frac{1}{2}}X = P$ (say) will give a set of n orthogonal factor measurements in unitary standard measure.²

Now what the psychologist wants to determine are the 'unknown functional relations' (as Pearson called them) between each of the observed measurements for traits or tests and the 'unobservable causes' or factors. These relations can best be expressed as correlations between each trait or test and the hypothetical factor in which we are interested: i.e. (to borrow Galton's metaphor), as the 'saturation' of each trait with the factor. These inferred correlations (F say) will be given by

$$F = MP' = LXP' = \Lambda\Lambda^{\frac{1}{2}}PP' = \Lambda\Lambda^{\frac{1}{2}}, \quad (\text{vii})$$

With this procedure we note that

$$M = LX = FP. \quad (\text{viii})$$

We have thus succeeded in *analysing* the observed trait-measurements, M , into two descriptive sets of figures— F , the factor-saturations, descriptive of the tests, and P , the factor-measurements, descriptive of the persons. Incidentally we may also observe that, with the principal axis procedure, not only the factor-measurements, but also the factor-saturations are themselves 'uncorrelated' (i.e., the product sums of the columns will be zero).³

As Pearson remarks, the essential data for resolving the observed measurements into their orthogonal components consist simply of the intercorrelations between those measurements and the squares of the standard deviations of those measurements (i.e., their 'variances,' as we should now say). From this point of view, it will be noted, the factorization of the correlation-table is a means

(*Mathematical Methods of Statistics*, 1946, p. 275; see also his diagram of the 'concentration ellipse' and three regression lines, p. 284; Pearson's paper (8) and his determinantal equation are quoted for the n -dimensional case, *ibid.*, p. 309). An interesting application is cited by Yule and Kendall, *Introduction to the Theory of Statistics*, p. 314, footnote 1: (the first edition also quoted Pearson's paper as describing the method of "fitting principal axes in the case of more than two variables," and gave his bivariate problem as a student's exercise: 1910, ed. pp. 333, 334).

¹ We should reach the same formulation if we started with the principle proposed by Edgeworth for estimating constants, which (as already noted) is virtually that of maximum likelihood (v. p. 103). The likelihood that a particular observational error, $d_{ij} = m_{ij} - x_i$ say, will occur in the attempt to measure i 's general factor ($i = 1, 2, \dots, N$) with the j th test will be

$$P(d_{ij} | x_i, p_j, \sigma_j, H) = \frac{\sqrt{p_j}}{\sigma_j \sqrt{2\pi}} \exp \left\{ -\frac{p_j(m_{ij} - x_i)^2}{2\sigma_j^2} \right\} dm_j;$$

and the joint likelihood of the whole system of such observational errors will be the product of $n \times N$ such functions. Hence, if all our tests have the same precision or weight (p), and if their standard deviations, σ , are all unity, we can take $x_i = \sum_j l_j m_{ij}$. The greatest likelihood will then be

attained when the direction cosines, l_j , are so chosen as to minimize the sum of the squares of the observational errors, $\sum_i \sum_j (m_{ij} - x_i)^2$.

² For a slightly more rigorous derivation of the essential formulae see Wilks, S. S., *Mathematical Statistics*, 1947, pp. 252–257. He emphasizes that "the theory of principal axes as discussed in this section includes no sampling theory," and does not require the assumption of a normal multivariate distribution. The sampling theory of the roots is summarized *ibid.*, pp. 250f.

³ In 1934 these *Notes on Factor Analysis*, originally prepared for research students, were compiled in the form of a memorandum at the request of the English Committee of the International Examinations Inquiry (of which I was a member), and were eventually published in abridged form as an appendix to their Report (see 23, p. 290, for the above proof; also 24, p. 78, equations 18 and 21). Much of the present paper is based on the fuller version of the notes, which are available in roneo'd copies. The idea of using vector and matrix notation had been suggested by Dr. W. F. Sheppard; but instead of his terminology and symbols (17, pp. 104f.), I adopted those proposed by Cullis (14). Roneo'd copies of the *Notes* are obtainable from the Laboratory.

to an end rather than an end in itself. Pearson, in fact, envisages the primary task as the determination of factor-measurements for the individuals, not of factor-saturations for the traits.¹

"To show that the methods can easily be applied to numerical problems," Pearson works out two imaginary problems involving two and three variables respectively. He admits that "the labour becomes more cumbersome if we have four or more characters which involve the determination of the greatest root and equations of the fourth or higher order." And, so far as I am aware, neither he nor Macdonell ever applied the full procedure either to the Scotland Yard data or to any other actual correlation tables. Instead certain approximate devices were proposed.

As a rough and tentative guide for determining the relative importance of the traits, Macdonell suggests calculating the square root of the minor of the self-correlation of the trait (i.e., in Pearson's notation, $\sqrt{R_{ii}}$, where R denotes the determinant of the correlations, and R_{ii} the minor obtained by crossing out the i th row and the i th column). This may be regarded as a plausible approximation to what the psychologist would call the saturation coefficient for the first or general factor.² Accordingly, for purposes of comparison I have appended at the foot of Table I the proportional values thus obtained for each trait by Macdonell's procedure.

III. BIPOLAR ANALYSIS WITH SELF-CORRELATIONS OF UNITY

(a) *Weighted Summation.* When the entire table of correlations is strictly 'synclinal' or 'hierarchical' (i.e., when it forms a square symmetric matrix of rank one), the calculations are relatively simple. In that case the self-correlations must also be proportional, and the factor-saturations are simply the square roots of these self-correlations (see 14, 1918, II, p. 134, eq. A). When they are taken as unity, and the table is no longer hierarchical, it can be progressively reduced to hierarchical form by simply multiplying it by itself, and repeating the process until the rows are as nearly proportional as seems desirable.³ In this way the factor-saturations for the first factor can be found; and, by using the same device with the residuals, the remaining factors can also be calculated without much difficulty.

This procedure, which I called 'table-by-table multiplication,' provides a fairly simple working-method for calculating the 'principal axis factors' as defined by Pearson's formulæ. It later appeared more accurate, as well as more expeditious, to substitute 'table-by-column multiplication'—a procedure which is now usually termed 'weighted summation' (23, pp. 467f.).

Table II gives the complete set of factor-saturations for Table I obtained in

¹ Here, to avoid unfamiliar circumlocutions, I have allowed myself to use terms which have come into vogue among psychologists since Pearson wrote his earlier articles. The above difference in the formulation of their initial task accounts, I fancy, for much of the subsequent controversy between educational psychologists, who commonly started with a problem similar to Pearson's and theoretical psychologists like Spearman, whose primary aim was to 'analyse the mind,' not to classify individuals.

² The value that Macdonell finally calculates is based on the expression $\sqrt{\frac{R}{R_{ii}}}$. This is equivalent to $\sigma_{i.12\dots n}$, if the traits are in standard measure. He then inverts the order. In the table the figures furnished by Macdonell's calculation have been multiplied by a constant so that the absolute values are of the same order as those furnished by other procedures.

³ This curious feature was noted when trying various forms of the 'intercolumnar criterion.' It may be recollected that Spearman and Hart proposed to test 'hierarchical order' (in Spearman's sense) by correlating the columns of the observed correlation table. To test 'hierarchical order' (in my sense) it was necessary to calculate what I called the 'unadjusted correlations'; and, to determine these, the computer must calculate the product-sums for every possible pair of columns: that, of course, is only another way of saying that he must multiply the matrix by itself. For a fuller explanation of the procedure I proposed, together with a worked example, see Thomson, 27, pp. 75-6. The method is really a corollary to what is known in matrix algebra as 'Sylvester's theorem.'

this way.¹ The accuracy with which Pearson's requirements have been fulfilled can be tested in two ways : first, it must be possible to reconstruct the $n \times n$ correlation table exactly from the n factors ; and secondly the $\frac{1}{2} n (n - 1)$ 'unadjusted correlations' between the n columns of factor-saturations must all be zero. The reader can easily satisfy himself that these conditions are met within the errors of rounding.

Of the seven factors thus obtained the first accounts for about one-half of the total variance, the second for one-fifth, the third for only one-tenth, and each of the rest for a twentieth or less. The first two are readily interpreted ; and, as we shall find in a moment, reappear with only slight variations in all the different analyses that can be offered for this table.

TABLE II. UNIT SELF-CORRELATIONS : (a) WEIGHTED SUMMATION
Pearson's Method of Principal Axes

Factor	I	II	III	IV	VI	VII	V	Square Sum
1. Head Length..	·5382	+·4462	+·7119	—·0519	—·0025	+·0051	—·0404	1·0000
2. Head Breadth	·4130	+·7836	—·2062	+·4117	—·0164	—·0059	+·0543	1·0000
3. Face Breadth..	·5751	+·6279	—·3087	—·4177	—·0170	+·0254	—·0659	1·0000
4. Foot ..	·8777	—·2184	—·0478	+·0309	+·4224	—·0489	—·0170	1·0000
5. Forearm ..	·8884	—·3390	—·0289	+·0691	—·1407	+·2645	—·0228	1·0000
6. Height ..	·8493	—·2197	—·0058	—·0567	—·1151	—·1187	+·4470	1·0000
7. Finger ..	·8529	—·2877	—·0569	+·0679	—·1527	—·1697	—·3602	1·0000
Square Sum ..	3·7994	1·5012	·6510	·3603	·2353	·1135	·3393	7·0000
Contribution to Variance (per cent.) ..	54·3	21·5	9·3	5·1	3·4	1·6	4·8	100·0

TABLE III. PEARSON'S METHOD APPLIED TO AN ARTIFICIAL CORRELATION TABLE
To illustrate the effect of treating Specific Factors as Common Factors

Factor	I	II	III	IV	VI	VII	V	Square Sum
1. Head Length..	·51	+·61	+·58	·00	·00	·00	·00	1·00
2. Head Breadth ..	·51	+·61	—·29	+·39	·00	·00	·00	1·00
3. Face Breadth ..	·51	+·61	—·29	—·39	·00	·00	·00	1·00
4. Foot ..	·84	—·29	·00	·00	+·45	·00	·00	1·00
5. Forearm ..	·84	—·29	·00	·00	—·15	+·42	·00	1·00
6. Height ..	·84	—·29	·00	·00	—·15	—·21	+·37	1·00
7. Finger ..	·84	—·29	·00	·00	—·15	—·21	—·37	1·00

¹ The factors are numbered in order of extraction, which is determined by the magnitude of their variances : factor V, however, is placed last to exhibit the factor-pattern more plainly. The saturations printed above differ somewhat from those given in earlier notes. For the tables circulated to illustrate my paper to the Royal Anthropological Society, the matrix was only squared three times, and the figures were calculated to three decimal places only ; a rough check was made by calculating the roots of Pearson's equation by a modification of the Newton-Raphson iterative method (suggested by W. F. Sheppard). However, the 'correlations' between the factor-saturations were not so near to zero as they should have been. To obtain the figures in Table II, a further series of successive approximations was undertaken, and the working carried to five decimal places. For these additional computations I am indebted to Dr. Banks. Macdonell, "by the laborious process of reduction," calculated the determinant of the entire correlation matrix, and found it to be 0·012129 (*loc. cit.*, p. 207) ; with the above analysis it is simply the product of the seven square-sums, viz., 0·012126. This close agreement testifies to the accuracy of the arithmetic. The reader will find several useful methods of calculation, based on successive approximation and other devices, in Duncan, Fraser, and Collar, *Elementary Matrices*, 1938, pp. 132f.

.I. The first yields a positive saturation for every trait ; it is plainly a 'general factor' for *body-size*. II. The second is a 'bipolar factor,' dividing the seven traits into two main groups—(i) those depending chiefly on the long bones, and (ii) those depending on the head and face ('limb-measurements' and 'head-measurements,'¹ as we may perhaps briefly call them). This second factor, therefore, might be regarded as a factor for '*head versus body-and-limbs*.' The five remaining factors show a peculiar distribution of positive, negative, and zero saturations : (the peculiarities are exemplified more clearly in Table III which will be explained in a moment : v. p. 112). If we sum the 'hierarchies' (i.e., rank-one matrices) obtained from these last five factors, we shall find that they contribute appreciable amounts to the variances in the leading diagonal, but comparatively little to the intercorrelations. This suggests that their chief function is to translate the unique or 'specific' factors, peculiar to each single test, into the form of bipolar 'common factors.'

As a mathematical device for transforming a correlation matrix to a more manageable shape, Pearson's procedure, which is merely what a mathematician would describe as "the reduction of a symmetric matrix to an orthogonal canonical form,"² is invaluable. But for the purposes of the psychologist it suffers from two serious defects, which must now be obvious : (1) from a practical standpoint, it is exceedingly laborious to apply when the number of correlated traits is much more than three ; (2) from a theoretical standpoint, it has the disadvantage of treating specific factors, which by definition should each be peculiar to a single trait, as common factors, shared by all the traits. In psychological measurements the specific factors, which must include the 'error-factors,' may be quite large. Hence for psychological work this treatment seemed particularly unsuitable. The modifications that I myself proposed sought to evade the two foregoing difficulties in the following way. First, by substituting simple summation for weighted summation, a short approximate method for practical computation can be provided : secondly, by substituting 'reduced self-correlations,' attributable to significant common factors only, for the perfect self-correlations of 1.00 assumed in Pearson's calculations, we can exclude specific factors from the very start. Let us consider each of these suggestions in turn.

(b) *Simple Summation*. It is a familiar fact that, unless the weights employed differ greatly in their numerical value, an unweighted sum of any set of measurements will always yield a good approximation to the weighted sum. Much labour will therefore be saved if we drop the differential weights, and simply add or subtract the trait-measurements (duly expressed in standard measure³) just as they stand. Adding the observed measurements will give measurements for the 'general factor' ; adding the residual measurements will give the supplementary factors.⁴

This is tantamount to reviving Galton's original suggestion. It is then easy to show that the correlation of each trait with the hypothetical factor, as thus estimated, will be proportional to the unweighted sum of the observed correlations (the self-correlations being still taken as unity throughout).

¹ It may be noted that, about this time, Pearson was specially interested in head-measurements. Galton (*Ann. Psych.*, V, 1898, pp. 245–298), Bain (*The Study of Character*, p. 195), and even Binet in his earlier work (*Ann. Psych.*, VII, 1900, pp. 314–429) believed that, if due allowance were made for body-size, mental and moral deficiency could be diagnosed by measuring the size of the head. Had they been right, the microcephalic imbecile would have a large negative measurement for this factor, and the genius a large positive measurement. Pearson subsequently showed that the correlation, though positive, was of no practical value : it proved in fact to be no larger than that between hair-colour and intelligence (*Biom.*, V, 1906, pp. 105–146).

² See, for example, McDuffee, C. C., *op. cit. sup.*, pp. 169f.

³ It is the unintentional weighting, introduced by *not* standardizing the measurements, that is most likely to lead to serious error. Galton's brief discussions, in (1) and elsewhere, seem to imply some such proof as is given below in the text.

⁴ For this approach to factor analysis, see 'Factor Analysis and Analysis of Variance, this *Journal*, I., pp. 6f.

Thus, take the first factor and the first trait. Let x_0 denote the factor-measurement, and $x_1, x_2, \dots, x_n$ the trait-measurements, all in unitary standard measure. Then, if $x_0 = x_1 + x_2 + \dots + x_n$,

$$r_{10} = \frac{\sum x_1 x_0}{\sqrt{\sum x_1^2} \sqrt{\sum x_0^2}} = \frac{\sum x_1 (x_1 + x_2 + \dots + x_n)}{\sqrt{\sum x_1^2} \sqrt{\sum (x_1 + x_2 + \dots + x_n)^2}} = \frac{\sum r_{ij}}{\sqrt{\sum \sum r_{ij}}} \quad (\text{ix})$$

To obtain the saturations, therefore, we have merely to add each column of correlations in Table I, and divide each sum by the square root of the grand total. The saturations for the supplementary factors can be obtained by treating the residual correlations in the same way.

The figures so reached are shown in Table IV. It will be seen that for the first two factors the pattern of signs is left unchanged, and that, with trifling exceptions, the relative size of the factor-saturations remains much the same as before, though the differences are somewhat diminished.

TABLE IV. UNIT SELF-CORRELATIONS : (b) SIMPLE SUMMATION

Factor	I	II	III	IV	V	VI	VII	Square Sum
1. Head Length	·6091	+·4349	+·5325	-·3952	·0000	·0000	·0000	1·0000
2. Head Breadth	·5327	+·6593	-·2484	+·1976	+·3222	-·2790	·0000	1·0000
3. Face Breadth	·6585	+·5141	-·2841	+·1976	-·3222	+·2790	·0000	1·0000
4. Foot ..	·8304	-·3649	-·1530	-·1851	+·1952	+·2164	+·1862	1·0000
5. Forearm ..	·8250	-·4700	+·0863	+·0978	-·1270	-·1753	+·1862	1·0000
6. Height ..	·8038	-·3598	+·2334	+·2974	+·1270	+·1753	-·1862	1·0000
7. Finger ..	·7984	-·4136	-·1667	-·2101	-·1952	-·2164	-·1862	1·0000
Square Sum ..	3·7421	1·5427	·5391	·4107	·3160	·3106	·1388	7·0000
Contribution to Variance (per cent.) ..	53·5	22·0	7·7	5·9	4·5	4·4	2·0	100·0

IV. BIPOLAR ANALYSIS WITH REDUCED SELF-CORRELATIONS

The second modification needs a fuller discussion, since, as we shall see, it can produce far more radical changes in the ensuing factor-pattern. Its aim is to avoid postulating a larger number of common factors than the significance tests will justify.

The first attempts¹ to calculate factors from complete correlation tables were those described in my article of 1909 (11). The fundamental purpose was to find evidence for what I called "the highest common factor," assumed to underlie representative measurements belonging to a given psychical field (namely, cognition), and determined from the entire table of data. Owing to the small numbers, the probable errors were large. But, within the margin permitted by these probable errors, and with certain notable exceptions, it seemed possible to fit the coefficients reasonably well by means of a single common factor only.

Now, if we apply Pearson's method to an ideal $n \times n$ table in which the intercorrelations are due to one factor only (like the theoretical tables for the hierarchy of kinship referred to above), we shall still obtain not one but n common factors; and if with Pearson we put the self-correlations equal to 1·00, the first of these n factors will *not* give a good fit to the intercorrelations. As we have seen, with a table of this kind the best fit is obtained by substituting figures for the self-correlations that conform with the proportions of the rows and columns.² In that case simple summation will yield a perfect

¹ The general notion had been put forward, with illustrations from Edgeworth, Pearson, and tests taken from Meumann, in two earlier papers, read to the Delian Society (1904) and the Oxford Philosophical Society (1906) respectively, but these efforts were not intended for publication.

² In referring to the working procedure proposed, Prof. Thomson observes (27, p. 25) "it is not quite clear how he (Burt) would have filled in the blank diagonal cells." In my earlier papers the method suggested was graphical interpolation. With tables like those described in the text, it was noted that, when the observed correlations are plotted, the lines joining the points for the several tests all tend to meet in a single point on the base line: (it was for this reason that, adapting a terminology employed by Pearson, such tables were called 'synclinal'). The appropriate figure for the self-correlation can then be found by noting "where the ordinate for each test met the line of slope."

fit with one factor alone ; and, even with empirical tables, a very close fit can often be secured with a single factor calculated in this way. It will be noted that inserting proportional figures in the diagonal is really equivalent to inserting only that portion of the total variance which is attributable to the single common factor : that is, it is tantamount to excluding the specific factor variance from the analysis at the very outset.

In Table I the coefficients are plainly not proportional throughout. Nevertheless, it can readily be shown that much the same argument still applies.

On these and other grounds, therefore, I ventured to contend that, instead of aiming at a *maximum* number of common factors, a safer policy is to aim at a *minimum*, retaining merely those that are fully significant—one general factor only, if that will suffice to account for the figures observed, adding others only when the evidence compels us. *Frustra sit per plura quod fieri potest per pauciora.*

In defence of this proposal, let me first draw the reader's attention to the artificial results of retaining the specific factors and including them in the variance to be analysed. The consequence will be seen most clearly if we apply Pearson's procedure to a fictitious correlation matrix in which the common factors are *known* to be very few, or even absent altogether.

(a) Let us take the extreme case first, and try factorizing a correlation matrix in which the inter-correlations are all zero. Table VA shows such a matrix ; and Tables VB and VC give the results of factorizing it first with differential weights, and then with equal weights, or, what amounts to the same thing, no weighting at all. It will at once be remarked that in Table VB the pattern of signs and zeros corresponds exactly to the pattern for the last four traits (traits 4 to 7) in Table II. (The factors to be compared are factors II, VI, VII and V : we are not concerned with factor I ; and factors III and IV drop out because all their saturations for the traits in question are numerically less than ± 0.07 and so non-significant, i.e., equivalent to zero.) Similarly the pattern of signs in Table VC corresponds to the pattern for the same last four traits in Table IV : (here, instead of dropping two factors, the pair III and IV and the pair V and VI can be merged into single factors on the ground that their sign pattern is the same).

TABLE V. THE EFFECT OF FACTORIZING A UNIT MATRIX OF SPECIFIC FACTORS

A. The Correlation Table					B. Factorized with Differential Weights				C. Factorized with Equal Weights			
Traits					Factors				Factors			
Traits	(1)	(2)	(3)	(4)	I	II	III	IV	I	II	III	IV
(1)	1.0	0.0	0.0	0.0	+.500	+.866	.000	.000	+.5	-.5	+.5	+.5
(2)	0.0	1.0	0.0	0.0	+.500	-.289	+.816	.000	+.5	+.5	-.5	+.5
(3)	0.0	0.0	1.0	0.0	+.500	-.289	-.408	+.707	+.5	+.5	+.5	-.5
(4)	0.0	0.0	0.0	1.0	+.500	-.289	-.408	-.707	+.5	-.5	-.5	-.5

(b) Now, instead of zero intercorrelations, let us take a case in which the intercorrelations are known to be due to two common factors only, and factorize the whole matrix as before, keeping 1.00 in the leading diagonal. Suppose, for example, that we start with the following factor-saturations for the seven body-measurements : (I) General factor—(i) for the four 'limb' traits, .80 ; (ii) for the three 'head' traits, .50 ; (II) Bipolar factor—(i) for the four 'limb' traits, .30 ; (ii) for the three 'head' traits, -.50. On reconstructing the table of correlations, we shall find that the inter-correlations for the limb-measurements are all .73, and for the head-measurements all .50, and that the cross-correlations between the limb and the head-measurements are .25.

With a table of this sort we should naturally assume that each of the seven traits had a unique or 'specific' factor of its own, and that this would bring each of the self-correlations or trait-variances up to unity. But, on factorizing our table by the foregoing procedure, we inevitably obtain seven common factors, and no specific factors, because the procedure does not envisage specific factors. The factor-pattern obtained is shown in Table III (p. 109). Now compare this with what we have already obtained from actual correlations (Table II). If, as before, we regard any empirical factor-saturation, whose numerical value is less than ± 0.070 , as statistically insignificant, and mentally substitute zero, we shall perceive a remarkable parallel between that table and Table III : the general pattern is almost exactly the same.

These results seem plainly to confirm the view that the last five factors, obtained from the empirical coefficients in Table I by applying Pearson's full procedure, can represent little else but specific factors, each really peculiar to a single test.

1. *With Correction for Specificity.* To meet these objections, we must introduce a further modification into the procedure. Our object is now to seek the smallest number of factors that will fit the observed intercorrelations within the margin of significance provided by the standard error; and to achieve this we have to insert in the leading diagonal, not the largest possible figures, namely, unity throughout, but the largest set of figures that can be consistently got with the small number of factors allowed to us.

So long as our working procedure assumed that the number of factors was the same as the number of tests (n say), the factor-measurements could be calculated exactly. But, if we are limited to only three common factors, supplemented by n specific factors, the common factor measurements will have to be estimates, not figures that are exactly calculable. Moreover, our aim is to weight the observed trait-measurements in such a way that the three hypothetical factors will contribute as much as possible to those estimates. Evidently all this will entail a different working formula.

An appropriate expression can be attained in various ways; and in previous articles I have indicated alternative lines of approach. As it happens, whichever argument is adopted, the equation eventually reached turns out to be the same. Instead of Pearson's equation (the 'characteristic equation' of the unreduced correlation matrix)

$$|R - \lambda I| = 0, \quad (\text{vi})$$

we find we must solve the analogous equation

$$|R_a - \lambda I| = 0, \quad (\text{x})$$

where

$$R_a = R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}}, \quad (\text{xi})$$

R_c denotes the 'reduced' correlation matrix, and R_u denotes the specific factors. It will be seen that the new matrix, R_a , is obtained by taking the 'reduced' correlation matrix, R_c , and then 'correcting it for specificity': (by the 'reduced correlation matrix' is merely meant the original matrix of observed correlations with 'reduced' variances inserted in the leading diagonal instead of unity). The saturations can then be obtained by reversing the effect of the 'correction': we have in fact

$$F_c = R_u^{\frac{1}{2}} L_a \Lambda_a^{\frac{1}{2}}. \quad (\text{xii})$$

The principle involved may be most readily understood if we consider the case where there is but one significant common factor, e.g., 'general intelligence': n tests will then yield n fallible and independent observations of a single variable. Now each of the tests may have a different 'reliability' or 'precision'; and, if the examinees are tested once only, this precision has to be estimated. The situation is very similar to that discussed in problems of 'least squares'; and, as we have seen, Gauss's original arguments and Laplace's principle of 'inverse' probability both lead to the device of 'preparing' the observation-equations by dividing each throughout by the square root of its unreliability: here therefore the divisor will be $r_{iu} = \sqrt{1 - h_i^2}$, where h_i denotes the reduced self-correlation for the i th test, due to the common factor.¹

Where there are several common factors, various working procedures may be adopted.² The results obtained are shown in Table VI. For the first two factors the sign pattern is unchanged; but the numerical magnitude of the saturations has

¹ It will be noted that this is quite different from Spearman's method of 'correcting for unreliability,' which really has a different purpose. I hope to deal with his criticisms in a later paper.

A somewhat similar procedure was proposed by Rhodes ('On Lines and Planes of Closest Fit,' *Phil. Mag.*, 7th ser., III, 1927, pp. 357-64). He endeavours to generalize Pearson's solution by taking into account (a) the fact that the standard deviations of the observed quantities, σ , may not be equal, (b) that further quantities, u , are required (measured in the same units) "to reduce them to a common base," and (c) that the standard errors involved in measuring each quantity may be different. For purposes of practical calculation, however, he finds himself eventually obliged to assume that the errors will be approximately equal, and that $u_i = \sigma_i$ throughout. In (23), pp. 316f., he suggests a slightly different procedure, and a method of successive approximation not unlike my own: as he points out, although "the arguments are different," the equation obtained is "the same in form as that used by Professor Burt in his memorandum."

² These are more fully described in roneo'd *Notes on Factor Analysis* (1934, sect. VI). The simplest is to factorize R_c (not R_a or R) and take trial-values for the columns of $F_c R_u^{-1}$.

TABLE VI. REDUCED SELF-CORRELATIONS: (1) WITH CORRECTION FOR SPECIFICITY

Factor	I	II	III	Square Sum
1. Head Length	·4323	·1924	·2347	·2785
2. Head Breadth	·3935	·5653	·5332	·7577
3. Face Breadth	·6418	·7301	—·0876	·9474
4. Foot	·8104	—·2078	·0184	·6997
5. Forearm	·8950	—·3909	·0053	·9539
6. Height	·7931	—·2215	·0115	·6777
7. Finger	·8192	—·2822	—·0211	·7507
Square Sum	3·5099	1·2148	·3409	5·0656
Contribution to Variance (per cent.) ..	50·1	17·3	4·9	—

undergone some alteration. The third factor is peculiar in showing fairly large positive saturations for the two head traits. This is perhaps due to the small saturations that these two traits have with the second factor.¹

With well-balanced tables the work is less laborious than might be supposed. But with the particular correlation table that we are analysing here the disadvantages of the method become quickly apparent in the course of the actual calculation. If we commence by conservatively assuming that two factors alone may be sufficient, and afterwards go on to try the effect of three or more factors, we find that, not only the size of the saturations, but even the sign pattern of the bipolar factor is apt to undergo drastic changes. Often the balance between antithetical traits—between those with positive and negative saturations respectively—may be quite upset. And generally, when the number in the sample is small and the collection of tests or traits not chosen according to a well-planned scheme, the specific factor variances are very much at the mercy of the level of significance that we decide to assume. With small samples they are often highly irregular; and, if the sample is enlarged, they may change so much as to produce an entirely different sign pattern. With certain tables the communality for one or more of the traits may even approach 1·00; and in such a case the traits in question would have a weight approaching infinity.

These are awkward results. It is scarcely plausible to suppose that the specific factor-variances could differ so widely as the calculations imply, much less that these differences could change so greatly with change in the size of the sample. In such cases, therefore, it seems difficult to justify the large differences in weighting that these variations entail.² For most purposes it would seem more reasonable to assume that, with a properly planned investigation and with a proper selection of traits, the specific factor-variances would show comparatively unimportant differences.

¹ Since most readers are more familiar with the saturations derived from reduced correlation tables without correction for specificity, I may here anticipate the comparison. The order of size for the saturations for the first factor as given in Table VI above, is the same as that obtained by weighted summation applied to the reduced correlation table without correction for specificity (Table VII); and the order for the second factor nearly the same as that obtained by simple summation in the ordinary way (Table VIII).

I should add that the figures for this method given in my earlier memorandum were based on nine iterations only, and did not yield a very close fit to the self-correlations assumed. My Research Assistant, Mr. A. R. Jonckheere, has very kindly continued the iterations until the discrepancies are less than $\pm 0\cdot003$. His figures have therefore been substituted in the table above.

² The fact is that, in psychological work, if we are to aim at a rigorous analysis, we need further evidence for estimating the effect of what I have called the error factor. This can only be done by repeating the tests or assessments, so as to obtain a set of correlations indicating their 'reliability,' as indeed was done in earlier investigations (cf. 11, 13).

If, therefore, in the absence of clear evidence to the contrary, we assume that, so far as correction is concerned, the error variances (R_{ii}) are approximately equal, then both the formula and the working procedure become greatly simplified. In that case, as a brief algebraic proof will readily show, the procedure reduces to an application of the ordinary method of weighted summation to the reduced correlation matrix, i.e., to the table of observed correlations taken just as they stand, with the self-correlations or 'communalities' in the leading diagonal. The same simplification also emerges in those cases in which we are justified in assuming only a single significant factor, namely, the so-called 'general' factor. When either of these assumptions is made, the laborious correction for specificity becomes unnecessary in calculating the saturations. It may, however, be plausibly argued that some such correction is still desirable before testing the residuals for significance; and, in fact, this correction was commonly made in my earliest investigations.

2. *Without Correction for Specificity.* (a) *Weighted Summation.* With the correlations shown in Table I, it is plainly impossible to secure a perfect fit with one factor alone. In general, provided we are free to insert appropriate common-factor variances, a covariance table for seven traits can always be expressed in terms of four factors, with no residuals. If, however, the covariance table takes the form of a correlation table, then it is further necessary that these variances shall be so chosen that none exceeds unity; and at times, particularly when the correlations are as large as those in Table I, this further requirement may compel us to invoke more factors than would otherwise be needed. Here five factors will be essential if the variances, or 'self-correlations' as they may be called, are not to exceed unity. Let us therefore take this as the minimum number, and examine the effect of fitting the table in such a way that each in turn shall account for a maximum amount of the total variance and that no residuals whatever shall be left when the analysis is completed.

There are various ways of discovering the requisite self-correlations or communalities; and (what is often overlooked) unless the appropriate number of traits is selected for the specified number of factors,¹ the self-correlations will not of necessity be uniquely determined. Here we can get sufficiently satisfactory figures for these constants after a few approximations, preferably with the simple summation procedure to be described in a moment. The calculation of the factor-saturations will then follow the familiar routine (26, pp. 467f.). The saturations finally obtained are shown in Table VII. It will be noted that we secure a perfect fit to the observed correlations with only five factors, and a very good fit with only two. The variances or their proportions (indicated at the foot of the table) can be used to provide approximate significance tests for the several factors.

(b) *Simple Summation.* On the same grounds as before, however, we may, for ordinary purposes, substitute unweighted summation for weighted summation. The working procedure has been fully described elsewhere (26, pp. 461f.). The results obtained with this further simplification are shown in Table VIII. The method of factor analysis thus finally reached is the one we have regularly used for nearly all practical purposes.²

¹ The number of factors must be one of the triangular numbers (1, 3, 5, 10, 15, ...) and the number of traits must be the next triangular number (i.e., for 1 factor 3 traits, for 3 factors 5 traits, and so on).

² It is only right to add that Professor Karl Pearson did not himself approve of these attempted modifications. At the same time I must record my deep indebtedness to him and to his staff for their suggestive criticisms and for their readiness to help both me and my fellow workers in our early efforts to apply his methods to educational and psychological data.

TABLE VII. REDUCED SELF-CORRELATIONS: (2) WITHOUT CORRECTION FOR SPECIFICITY. (a) WEIGHTED SUMMATION

Factor	I	II	III	IV	V	Square Sum
1. Head Length	·4435	+·3065	+·0386	—·0376	·0782	·2998
2. Head Breadth	·3792	+·7823	—·0420	+·0970	—·0094	·7675
3. Face Breadth	·5101	+·5501	—·0090	—·0653	—·0359	·5688
4. Foot	·8453	—·1338	—·0032	—·1075	+·0146	·7440
5. Forearm	·9229	—·3054	—·0658	+·2225	+·0054	·9990
6. Height	·8452	—·1546	+·3471	—·0206	—·0210	·8595
7. Finger	·8486	—·2210	—·2670	—·0990	—·0150	·8598
Square Sum	3·6015	1·1925	·1995	·0865	·0084	5·0884
Contribution to Variance (per cent.)	51·5	17·0	2·9	1·2	0·1	72·7

TABLE VIII. REDUCED SELF-CORRELATIONS: (2) WITHOUT CORRECTION FOR SPECIFICITY. (b) SIMPLE SUMMATION

Factor	I	II	III	IV	V	Square Sum
1. Head Length	·4894	+·2280	+·0586	—·0171	+·0681	·2998
2. Head Breadth	·5060	+·7065	—·0846	+·0697	—·0160	·7675
3. Face Breadth	·5959	+·4555	+·0260	—·0526	—·0521	·5688
Total	1·5913	—1·3900	·0000	·0000	·0000	—
4. Foot	·8107	—·2749	+·0498	—·0935	+·0064	·7440
5. Forearm	·8574	—·4570	—·1740	+·1525	+·0384	·9990
6. Height	·8067	—·2988	+·3036	+·1632	—·0224	·8595
7. Finger	·7991	—·3593	—·1794	—·2222	—·0224	·8598
Total	3·2739	1·3900	·0000	·0000	·0000	—
Square Sum	3·5323	1·2614	·1686	·1160	·0101	5·0884
Contribution to Variance (per cent.)	50·5	18·0	2·4	1·7	0·1	72·7

Comparison of the Bipolar Procedures. Let us now briefly review the results obtained by the different methods so far described. There are two main questions to be borne in mind. First, is the method proposed by Pearson (or some derivative of it) really applicable to mental characteristics? Spearman from the very outset maintained that it was not. Secondly, do the proposed simplifications lead to such different results that they “evidently defeat the original intention of the proposal and destroy its mathematical accuracy”? Pearson held that they do.

(i) *Contributions to Variance.* As regards the results, the chief alterations are due to the use of reduced self-correlations. When these are substituted, nearly 30 per cent. of the variance, which with the older procedure would be spread over the seven common factors, is assigned to specific factors. Now in anthropometric work I venture to suggest that what we have since learnt about physical growth, and particularly about the importance of ‘heteronomic’ as distinct from ‘isonomic’ growth-tendencies, is fully in keeping with the implied assumptions;¹ and in psychological work the unreliability of our assessments must involve large specific factors.

¹ See Burt, C., ‘The Factorial Study of Physical Types,’ *Man*, LXXII., 1944, p. 85 and refs.

Even with this small battery, the change affects the main factors less than might be expected. With every method so far employed, between 51 and 54 per cent. is attributed to the first or 'general' factor; between 17 and 22 per cent. to the second; and less than 10 per cent. to the third. These are very much the same proportions as we get on factorizing mental characteristics (cf. 23, p. 358). Here, it is true, the contribution of the first bipolar factor is somewhat larger than that usually found in early work on mental tests.¹ But the difference is mainly due to the way the traits are chosen, not to the fact that one set is physical and the other mental.

(ii) *Sizes and Signs of Saturations.* For the first two factors, the pattern of signs remains the same with all the methods; and the size of the saturations follows much the same order. For the third factor the chief points of disagreement relate to the measurements for Height and Head and Face—the only traits that provide appreciable saturations. But here the differences depend on the way in which each procedure is compelled to translate the specific factors into terms of common factors. So far as the significant factors are concerned, the close agreement between the two last methods (weighted summation and simple summation) is particularly striking. Thus, the neglect of any differential weighting (the point in Galton's procedure which Pearson most strongly criticized) here leads to comparatively small changes, and the saving in labour is immense.

These then are the main justifications for the modified procedures I ventured to propose; and I am tempted to maintain that the whole trend of factorial research during the past forty years, in the field both of physical and of mental measurement, has fully borne out the suggestion that Pearson's principles might be profitably applied to psychometric data as well as to anthropometric, and has in the main confirmed the modifications tentatively put forward.

V. GROUP FACTOR ANALYSIS

(a) *Non-overlapping Factors.* With all the foregoing modes of analysis one invariable feature has been the large bipolar factor, drawing a sharp contrast between the head-measurements (as we have called them) and the limb-measurements. If we tried to give this factor a concrete causal meaning, we should have to suppose that it represents some genetic or metabolic condition (possibly quite complex) which, when acting positively, causes an individual to develop a head that is disproportionately large as compared with his limbs, and, when acting negatively, causes him to develop limbs that are disproportionately large as compared with his head. Now, when we turn to results obtained with mental tests, most psychologists would find it somewhat artificial to think of a condition or capacity that could act in two opposite ways. Even in the field of physical growth many would probably prefer to postulate (in addition to the general growth factor) two *separate* tendencies, each having a positive effect only—one making for head-growth, the other for limb-growth.

We are therefore led to ask whether it is possible to analyse our correlation table in such a way that the factors will have saturations that are solely positive. Now this suggestion is strongly reinforced by two well-marked features which Macdonell's table happens to present. First as the reader will have noticed it falls into four distinct quadrants. In the north-east quadrant the coefficients range from .39 to .62; in the south-west from .66 to .79; in the north-west and south-east quadrants from .13 to .36. In fact, as Pearson noted, we are confronted with correlations of two distinct levels, high figures for the intercorrelations between 'like organs' and low figures for the cross-correlations between 'unlike organs.'² But secondly there is

¹ That is why this particular correlation table failed so conspicuously to conform with the inter-columnar and tetrad-difference criteria. It was this failure that formed Spearman's strongest argument against my proposal. But, when tests of verbal and non-verbal abilities, or assessments of extraversive and introversive traits, are included in the same battery, the first bipolar becomes just as conspicuous. His other arguments against determining factors from the *entire* correlation table I shall deal with in a later paper.

² Cf. *Grammar of Science*, pp. 402f. Palin Elderton (*Primer of Statistics*, 1909, p. 70), quoting figures chiefly from the same table, suggests recognizing three levels—'medium, high, and low correlation.'

another peculiarity that may easily pass unnoticed. If the reader will examine the cross-correlations in the north-west (or south-east) quadrant he will see that the rows and the columns are *almost exactly proportional*: the figures in fact here approximate very closely to a perfect 'hierarchy' or 'matrix of rank one.' It follows that these twelve cross-correlations could readily be accounted for by a single factor only.¹ And, if one will serve, why invoke two or more?

Accordingly, as I ventured to point out in the paper cited above (p. 109), it is tempting to suggest that Macdonell's table might be analysed into uncorrelated factors in a way quite different from that proposed by Pearson. We could in fact regard it as built up by a process of simple superposition, as consisting, that is, of a low set of correlations due to some basic factor pervading the whole table, and two sets of supplementary correlations (due to limited group factors) which have been, as it were, added on to these low values, and so produce two square clusters of high coefficients, straddling the main diagonal. Indeed, this mode of analysis would actually be more in accordance with Pearson's doctrine of 'like' and 'unlike' organs.

Still keeping to the 'vector law' ($r_{xy} = r_{xb}r_{yb} + r_{xg}r_{yg}$, where b and g denote basic and group factors, respectively), we can easily test this suggestion. We begin by calculating and eliminating the effects of the basic factor, which alone is assumed to cause the correlation between unlike organs. For this purpose we apply the formula for simple summation (suitably adapted²) to the oblong submatrix of cross-correlations. The saturations so obtained are shown in the first column of Table IX. We subtract the effects of this first factor from the intercorrelations in the two square submatrices, and then go on to compute the saturations for the two narrower factors that enter into one or other of these limited groups of traits. These supplementary saturations are shown in the second and third columns of the table.³

TABLE IX. GROUP FACTORS: (a) NON-OVERLAPPING

Factor						I	II	III	Square Sum
1. Head Length	..	..	..	..	..	·4948	·2838	—	·3253
2. Head Breadth	..	..	..	..	..	·2598	·9620	—	·9929
3. Face Breadth	..	..	..	..	..	·5076	·5051	—	·5128
Total	..	..	..	..	..	1·2622	1·7509	—	—
4. Foot	..	..	..	..	..	·7195	—	·5014	·7691
5. Forearm	..	..	..	..	..	·5778	—	·8161	1·0000
6. Height	..	..	..	..	..	·6879	—	·4636	·6881
7. Finger	..	..	..	..	..	·6117	—	·5816	·7125
Total	..	..	..	..	..	2·5969	—	2·3627	—
Square Sum	..	..	..	..	..	2·2690	1·4707	1·2610	5·0007
Contribution to Variance (per cent.)	..	..	..	..	..	32·4	21·0	18·0	71·4

¹ This feature was also noted in the case of mental tests or assessments, where the abilities measured fell into two or three sharply demarcated groups: e.g., with the tests of visual, audile, and kinæsthetic imagery used by J. A. Davies and myself (cf. *J. Exp. Ped.*, I, 1912, p. 251; cf. also 13, Table XXII, p. 61). For working methods see *J. Psych.*, VI, 1938, pp. 339f. (reprinted from *Notes on Factor Analysis*, 1934).

² Since the submatrix is not symmetric, we can no longer use the square root of the grand total as the divisor for the row- or column-sums: if we did, the saturations obtained would not fit the rest of the table. Two divisors are now needed, one for column-sums and the other for row-sums. To determine the proportion between the two divisors the following principle provides a simple working procedure. If the 'basic factor' is to represent the same kind of component as the 'general factor,' the proportions of the saturations should be approximately the same. Hence the proportions of the two corresponding sets of subtotals for those saturations, and therefore of the divisors, should also be approximately the same. This is *strictly* true only when certain relations hold between the basic and group-factor saturations; but in most cases it provides a convenient starting point, which can be corrected later by successive approximation if required: (see *Notes* already cited).

³ This procedure must be carefully distinguished from that later proposed by Spearman. With a battery of (i) verbal and (ii) perceptual or spatial tests, he takes the former group as 'reference-values,' i.e., as pure tests of g , and looks for a group factor solely in the latter. My determination of the general factor is based on the *whole* correlation table, and I then look for *two* (or preferably more than two) group factors.

(b) *Overlapping Factors*. So far we have assumed that each group factor will be confined to its own sub-matrix. But the possibility of an overlap cannot be excluded.

Racial considerations, for example, might suggest that the factor making for tall stature and long limbs should show a positive saturation for length of head and a negative saturation for breadth of head, since dolichocephalic persons tend to be taller and more slender than brachycephalic. The largest of the residuals left by the method just described bears out this suggestion ; and it is further confirmed by the peculiarities already noticed in the bipolar analysis.

To allow for such an overlap, the most obvious procedure is to employ the method of rotating axes first introduced by Garnett.¹ If we turn back to the factors obtained by the bipolar analysis, we can apply one of the many devices current in matrix algebra for suppressing the larger negative saturations, and so construct a suitable rotation matrix. But probably the simplest procedure is to reconstruct a new set of correlations from the non-overlapping group-factor matrix, analyse this new table by the bipolar method, and then compute the transformation matrix required to convert one set of factors into the other. The transformation matrix so obtained will be *exactly* orthogonal. We can accordingly use it to postmultiply the first three columns of significant factor-saturations shown in Table VII. We thus obtain the saturations for the overlapping set of group factors shown in Table X.

TABLE X. GROUP FACTORS : (b) OVERLAPPING

Factor					I	II	III	Square Sum
1. Head Length	..	..	..	..	·4200	·3368	·0187	·2949
2. Head Breadth	..	..	..	..	·3338	·8018	—·0900	·7624
3. Face Breadth	..	..	..	..	·4792	·5774	·0713	·5633
Total	..	..	..	..	1·2330	1·7160	·0000	—
4. Foot	..	..	..	..	·6863	·0360	·5129	·7354
5. Forearm	..	..	..	..	·5946	—·0021	·7879	·9743
6. Height	..	..	..	..	·8359	—·0945	·3532	·8323
7. Finger	..	..	..	..	·5419	·0606	·7089	·7999
Total	..	..	..	..	2·6587	·0000	2·3629	—
Square Sum	..	..	..	..	2·3343	1·5246	1·1036	4·9625
Contribution to Variance (per cent.)	..	..	..	..	33·3	21·8	15·8	70·9

Now let us compare these rotated saturations with those furnished by the more direct procedures. With the method we have employed the saturations for the overlapping three group factors will yield precisely the same reconstructed correlations as would be obtained from the first three factors of the bipolar analysis ; and by dropping all factors except these three we have only lost $(72·7 - 70·9) = 1·8$ per cent. of the variance. The modifications introduced by the overlapping are comparatively slight. The relative size of the factor-saturations has undergone little alteration ; and the few conspicuous instances of overlap could readily be accounted for on racial or endocrinological grounds.

Finally, it should be observed that, since the general-and-bipolar factors were orthogonal, the new overlapping group factors must be orthogonal too. If then our aim was to reduce the seven trait-measurements to the smallest number of significant but uncorrelated factors, having (so far as possible) positive saturations only, this last table would provide the best result of all. But I should maintain that a preliminary bipolar analysis is nearly always advisable to determine the main lines of division.

¹ 15, 1919, pp. 108f. ; cf. also *Education and World Citizenship*, 1921, pp. 122 and Appendix B.

VI. SUMMARY AND CONCLUSIONS

1. The paper attempts to illustrate, by a concrete example, the similarities and differences between the results of the commoner methods of factor analysis. Its chief purpose is to show how all these procedures, directly or indirectly, have developed out of, or are related to, the method of 'principal axes' suggested by Pearson in 1901. The example chosen is the table of correlations between physical measurements, calculated under Pearson's direction, and put forward by him as a suitable case for this mode of analysis.

2. *Bipolar Analysis*: (a) *With Unreduced Self-correlations*. Pearson's method is first applied in its original form. This is equivalent to what is now known as 'weighted summation,' with 'communalities' of unity. It seeks to obtain a maximum number of factors, namely, n factors for n traits. It is concluded that this version is not the most appropriate, at any rate for psychological work.

(b) *With Reduced Self-correlations*. A safer plan is to assume only the minimum number of factors required to account for the observed intercorrelations, or the minimum number of factors that are statistically significant. This entails substituting 'reduced' self-correlations in the initial correlation table, the rest of the variance being assigned to specific factors.

(i) In theory, as has elsewhere been shown, this reduction leads to a modification of Pearson's formula and a working procedure which is even more elaborate, viz., one involving a correction for specificity. In practice, however, this modification possesses certain disadvantages, which in turn require further simplifications.

(ii) A simpler formula can be derived by assuming that, so far as correction is concerned, the specific factor-variances can be treated as approximately equal. This is equivalent to applying Pearson's method to the reduced correlation table just as it stands, i.e., without explicit correction for specificity.

(iii) It is maintained that the labour still required would only be justified with a highly accurate set of measurements, and that in psychology the method of simple or unweighted summation may, as a rule, be substituted for that of weighted summation.

3. *Group Factor Analysis*. With certain tables, including that studied here, there is reason to believe that simpler and more intelligible results can be obtained by seeking group factors with positive or zero saturations instead of bipolar factors with positive and negative saturations. This leads to analyses by (a) non-overlapping, and (b) overlapping group factors. The lines of division for the former are based on the preliminary bipolar analysis; and the results in turn yield an orthogonal transformation matrix for converting the bipolar factors into overlapping group factors.

4. The factors obtained by these different methods are compared in detail; and it is maintained that the results of recent research justify the various suggestions made, and provide a sufficient reply to the objections raised by Pearson and Spearman.

REFERENCES

1. Galton, F. (1888). 'On correlations and their measurement,' *Proc. Roy. Soc.*, XLV, 135f.
2. Galton, F. (1889). *Natural Inheritance*. London: Macmillan.
3. Edgeworth, F. Y. (1892). 'On correlated averages,' *Phil. Mag.*, XXXIV, 190f.

4. Pearson, K. (1896). 'Mathematical contributions to the theory of evolution,' III, *Phil. Trans. A*, CLXXXVII, 253-318.
5. Edgeworth, F. Y. (1896). 'Supplementary notes on statistics,' *J. Roy. Stat. Soc.*, LIX, 529-539.
6. Yule, G. U. (1897). 'On the theory of correlation,' *J. Roy. Stat. Soc.*, LX, 812-824.
7. Pearson, K. (1900). *The Grammar of Science*. London : A. & C. Black.
8. Pearson, K. (1901). 'On lines and planes of closest fit to systems of points in space,' *Phil. Mag.*, 6th ser., 559f.
9. Macdonell, W. R. (1902). 'On criminal anthropometry and the identification of criminals,' *Biometrika*, I, 177-227.
10. Yule, G. U. (1907). 'On the theory of correlation for any number of variables,' *Proc. Roy. Soc.*, A, LXXIX, 182-193.
11. Burt, C. (1909). 'Experimental tests of general intelligence,' *Brit. J. Psych.*, III, 94-176.
12. Brown, W. (1911). *The Essentials of Mental Measurement*. Cambridge : University Press.
13. Burt, C. (1917). *The Distribution and Relation of Educational Abilities*. L.C.C. Report. London : King.
14. Cullis, C. E. (1918). *Matrices and Determinoids*. Cambridge : University Press.
15. Garnett, J. C. M. (1919). 'On certain independent factors in mental measurements,' *Proc. Roy. Soc.*, A, XCVI, 91-111.
16. Pearson, K. (1920). 'Notes on the history of correlation,' *Biometrika*, XIII, 25-45.
17. Sheppard, W. F. (1923). *From Determinant to Tensor*. Oxford : Clarendon Press.
18. Pearson, K. (1924). *Life of Galton*. Cambridge : University Press.
19. Board of Education (1924). *Report on Psychological Tests of Educable Capacity*. London : H.M. Stationery Office.
20. Burt, C. (1926). *The Measurement of Mental Capacities*. Edinburgh : Oliver and Boyd.
21. Rhodes, E. C. (1927). 'On lines and planes of closest fit,' *Phil. Mag.*, 7th ser., III, 357-364.
22. Dent, Beryl (1935). 'On observations of points connected by a linear relation,' *Proc. Physical Soc.*, XLVII, 92-106.
23. Burt, C. (1936). 'The analysis of examination marks : a review of methods of factor analysis in psychology' (ap. Hartog, P., Rhodes, E. C., and Burt, C., *Marks of Examiners* : also as a separate pamphlet). London : Macmillan.
24. Burt, C. (1937). 'Correlations between persons,' *Brit. J. Psych.*, XXVIII, 59-95.
25. Roos, C. F. (1937). 'A general invariant fit for lines and planes where all the variates are subject to error,' *Metron*, XIII, 1-20.
26. Burt, C. (1940). *The Factors of the Mind*. London : University of London Press.
27. Thomson, G. H. (1948). *The Factorial Analysis of Human Ability*. Third Edition. London : University of London Press.

BOOK REVIEW

Cybernetics : or Control and Communication in the Animal and the Machine. By Norbert Wiener. Hermann et Cie., Paris. 1948. Pp. 7-194. fr. 820.

This book has been impatiently awaited by those who were fortunate enough to have heard Professor Wiener discuss his theories when he visited this country two years ago. It is difficult to give a short account of the work that will yet do justice to the profusion of ideas with which it is filled.

In an introductory chapter the author tells how his interest in computing machines and predictors led him to consider some of the problems of the central nervous system. In all such systems a message, which may be mixed up with noise, is fed into the apparatus ; and this has then to interpret the message and act accordingly. Under such conditions the theory of operation reduces to problems in time series. Other scientists were also confronted with similar questions of control and communication, sometimes in animals, sometimes in machines ; and together they formed a group to discuss their common problems. To avoid bias towards a particular subject-matter, they decided to call this basic investigation 'cybernetics,' the science of steersmanship. The theory is a general theory of servomechanisms, which are concerned with making the 'output' pattern of a system conform to the 'input' pattern. In dealing with it the author has decided to "confine his efforts to those fields, such as physiology and psychology, most remote from war and exploitation."

The first chapter points out that a Newtonian system is symmetrical with respect to time ; its fundamental laws are unaltered if the time variable, t , is replaced by its negative. In biology this reversibility no longer holds ; both the individual and the race are "like arrows pointed in one direction, from the past to the future." Moreover, the reduction of thermodynamics to statistical mechanics—in which not a single system, but a statistical distribution of systems, is considered—has led away from the reversible time of Newton to the irreversible time of Willard Gibbs. When we attempt to discuss biological systems from a mathematical point of view, energy is no longer an adequate concept upon which to build a theory. We ought rather to consider "the information supplied" (e.g., to the nervous system). Nor are we concerned with the behaviour of the system for a *single* input, but have rather to consider it as a mechanism that must give satisfactory performance for a whole class of inputs. Hence the theory must be a statistical theory.

The next three chapters, which are predominantly mathematical, are possibly the most valuable in the book. The fundamental theory is lucidly expounded. The technique involves the use of Haar measure, the Lebesgue-Stieltjes integral, and that generalized harmonic analysis which formed the subject of a classical paper in *Acta Mathematica*, and was enlarged upon by the author in collaboration with Paley. He begins by outlining the theory of groups, and gives a short but remarkably clear account of Birkhoff's ergodic theorem and its extension by Koopman and von Neumann. He then points out that any message may be treated as a time series with an associated quantity of information, which for a

curve $f(x)$ is defined as $\int_{-\infty}^{\infty} [\log_2 f(x)] f(x) dx$. This is the negative of a quantity usually defined as entropy ; and, on the average, has properties usually associated with entropy : e.g., processes that lose information are analogous to those that gain entropy ; no operation on a message can, on the average, gain information.

It is shown that if a message, homogeneous in time, is a function (or a set of functions) of time, and if it forms one of an ensemble of such sets with a well-defined probability distribution, unchanged by the change of t into $t+\tau$, then the ergodic theorem may be applied. Consequently, in the case of such an ensemble, we can deduce the average value of any of its statistical parameters from the record of any of the component time series by using a time average instead of a phase average. This principle is then applied to the design of predictors and wave filters.

The concepts of feed-back and oscillation are discussed in connexion with cerebellar ataxia (in which the purposive movement of a limb ends in violent oscillation). The conditions under which a feed-back mechanism will go into oscillation are described at length; 'homeostasis' is taken as an example of non-neural feed-back.

Chapter V expands the analogy between computing machines and the nervous system. By coding the data in the binary system, the structure of a computing machine can be considered as a set of relays, capable of 'on' and 'off' positions, dictated by the results of previous operations. Neurons may be considered as relays; and a 'memory' may be produced by having impulses travel round closed circuits, or by a set of condensers charged and 'scanned' by a pencil of cathode rays forming one of the leads to the condensers. Conditioned reflexes are then described in terms of an 'affective tone totalizer,' which would act as a feed-back mechanism. The impossibility of analysing 'information' in terms of matter and energy is again stressed; 'information' forms a category of its own.

The problem of stimulus equivalents is dealt with in some detail. A square is recognized, no matter what its size or position in the visual field. The possible perspective transformations form a 'group' which defines several sub-groups of transformations, e.g., translation and rotation. We can, for example, consider a process of group scanning, whereby a fixed comparison region is compared with some other region. If, at any stage of the scanning, the image of the region, which is to be compared under some one of the transformations scanned, happens to coincide more perfectly with the fixed pattern than a given tolerance allows, then the two regions are said to be alike. This theory is applied to the visual cortex; and a diagram is given to show a concrete form of the system. This schematization, it is claimed, suggests the fourth layer of the cortex. The alpha rhythm is associated with form perception, and is a sweep rhythm similar to 'scanning' in a television set. When we look at something, the rhythm is obscured, and then acts as a carrier for other rhythms.

The analogy with the computing machine is used for certain problems of psychopathology. Functional nervous disorders are said to be disorders of memory, of the circulating 'information.' When a computing machine starts to make mistakes, the computer clears it; and this operation is likened to sleep in a human being. If the machine still works badly, the computer can feed in an abnormally large electrical impulse in the hope of interrupting the false cycle; this is compared to shock therapy. Again, there is a limitation to the size of a telephone exchange built to a fixed plan; and such a limitation would apply to the brain. Overloading would cause a jam. The superiority of the human brain, it is argued, depends on the length of the neuron chains: the higher the process involved, the longer the chain. Such long chains are said to be especially sensitive to overloading.

It must be admitted that the evidence for all these statements is not substantial; and would seem merely to justify the statement that the human brain is highly complicated, and consequently easy to disorganize. The problem of the dominant hemisphere is discussed in similar terms. The last chapter attempts a general and personal discussion of some problems in sociology and economics; and the author concludes that much of the subject-matter of these sciences cannot be treated by the scientific method.

The subject-matter of the book, it will be seen, really falls into two parts, the mathematical discussion of fundamental theory and the application of this theory to problems mainly of neurophysiology. The mathematical development is as lucid and masterly as one would expect; and the reviewer is glad to see that the author is convinced that the only adequate treatment for biological problems of this kind is in terms of statistical distributions, and that any attempt to formulate them in terms of Newtonian mechanics is futile. It cannot be doubted by any who have been in contact with the actual biological material that these questions demand the use of the most modern methods, be they those of Professor Wiener, integro-differential equations, or stochastic process theory.

But when the author begins to discuss the operation of the central nervous system, the clarity of the earlier chapters seems to be lost. Part of the value of a mathematical theory lies in its ability to describe the material concisely and to predict the results to be gained from experiments. Professor Wiener gives no indication as to how this could be carried out; in fact, almost all his discussion of the operation of the nervous system could have been

carried out without the mathematical theory. As a model, the computing machine may show how a nervous system could be constructed ; but there is little to show that it is, in fact, constructed in this way. Take a particular case : the mode of operation of the striate area is explained in terms of an analogy with photo-cells and oscillators ; and it is said that a leading anatomist asked whether the diagram represented the visual cortex. The reviewer cannot see the remotest resemblance between the two. Reference to the paper of McCulloch and Pitts shows a more complicated, but still highly schematic, diagram ; but no adequate reasons are given for thinking that this area does operate in the manner suggested. It is not very difficult to construct a number of different circuit diagrams with just as much plausibility.

Once more, it is suggested that many disturbances of the nervous system are due to the overloading of the longest chains of neurons, and that these long chains are ' responsible ' for ' the processes which are recognized to be the highest in our scale of valuation.' No convincing evidence is presented for these statements ; and throughout the later chapters the discussion rests upon supposed analogies. Metaphor may be the soul of poetry, but can become an incubus to science.

These criticisms may seem pedantic. But neuroanatomists and neurophysiologists, to say nothing of psychiatrists, will find grounds for many similar criticisms in the latter part of the book. One is left with the feeling that it contains numerous generalizations quite unsupported by a critical examination of the data. To convince the sceptic, a mathematical theory must be shown not only to apply to the data generally, but specific examples of its application should be given.

This review, however, must not end on a critical note. Professor Wiener's book embodies the first mathematical theory capable of discussing the central nervous system at all adequately ; and all who read with the attention worthy of the work will find new horizons in place of fog. It is written in a charming and informal style ; and is full of *obiter dicta* well worthy of quotation. There is room only for one, which may perhaps console those who consider that the theory of probability affords too vague and easy a way out of all difficulties : ' Chance is as relentless a mistress as Force.'

Donald Sholl

CORRESPONDENCE

To the Editors, British Journal of Psychology: Statistical Section.

Sirs,—The paper by Rao and Slater in the last issue is valuable, not only because it draws attention to a useful technique that is not so well known as it should be, but also because it is put in a form that exemplifies the computational procedures. Unfortunately, it is marred by two errors that will seriously hinder those who use it for the latter purpose. (1) On p. 20, the formula for the Δ statistic is correctly given ; but the degrees of freedom of the numerator are stated to be those of Table IV (the Within-Groups matrix). This is wrong. They should be the degrees of freedom of Table III (the Between-Groups matrix). (2) At the top of p. 26, the method of obtaining the variance of the canonical variates is described. It should be derived from Table VII, the Within-Groups Dispersion matrix, not from Table IV, the Product-Sum matrix.

Two minor criticisms can be made. The values for the canonical variates are given in absolute scores. Should they not be given as deviants from the general mean, which is statistically better, or as deviants from the normal group, which might be psychologically clearer ? Finally, since this *Journal* has regularly used matrix notation in its pages, it would be helpful to readers if authors would continue to do so. Thus the formula on p. 21 would become

$$D^2 = d' A_{ij}^{-1} d'$$

and that on p. 27 would be much clearer. MAX HAMILTON.

SIMPLE STRUCTURE : A CRITICAL EXAMINATION

By H. A. REYBURN and M. J. RAATH

The University, Cape Town

I. The Claims of Simple Structure. II. Simple Structure and Parsimony. III. Simple Structure and Invariance. IV. The Problem of Identity of Factors. V. A Consideration of Some of Cattell's Results. VI. Summary.

I. THE CLAIMS OF SIMPLE STRUCTURE

When the conception of simple structure was put forward as a guide to the rotation of axes in factor analysis, it was commended on three grounds. (a) It provides a definite criterion, for, it was contended, simple structure is unique : its conditions can be satisfied only in one way, and so the investigator knows when he has gained his objective. (b) Invariance of factor structure can be secured : each variable will be explained by the same factor when transferred from one battery to another. (c) It is in accordance with the principle of parsimony, in that it simplifies the analysis of the variables by assigning fewer factors to each. Moreover, it now seems to be assumed that the scheme of factors which simple structure implies is a probable hypothesis in itself when applied to human personality. These various points will be considered in turn.

II. SIMPLE STRUCTURE AND PARSIMONY

First, however, it is desirable to clear out of the way a possible misconception. The use of oblique axes for purposes of simple structure is legitimate under certain conditions and in specific problems. But it complicates matters by requiring one or more additional factors to account for their own interrelations. These 'secondary' factors are necessary to explain the data ; and in that sense they are as real as any so-called 'primary' factors. Further, it is obvious that an analysis which produces a number of 'primary' factors to relate the data in the first instance, and one or more 'secondary' factors to relate the 'primary' factors, is not, as a whole, a simple structure. *Pure* 'simple structure', if the term may be allowed, is orthogonal ; for only so can all the factors conform to the requirements which simple structure implies.

Now if we interpret simple structure in this rigorous way, it does not provide a likely hypothesis for human nature. Applied to the growth of the physical organism it seems out of the question : for the factors which differentiate growth in different organs imply as their basis some deeper unifying process in order that there may be an organism at all. Again, it is hardly possible to maintain that intellectual performance is the result of entirely independent factors, none of which limit or react on the others. There must be a mind before it can become specialized : and if it is described in terms which *prima facie* exclude a general factor, the factors admitted must interlace and interlock in some manner which brings one or more general factors back again. Nor is a pure simple structure a likely hypothesis in the field of temperament and personality. One or more general factors seem essential to any understanding of behaviour ; and if such factors are excluded in the first instance, they must be restored later : that is, if simple structure should be found in an analysis of personality traits as a whole, it must be oblique, and will require a further analysis to bring out the other underlying unities.

This brings us to a further point, the parsimony of oblique simple structure analysis when applied to personality as a whole. The position here is plain : simple structure does not reduce the total number of descriptive factors, but it does reduce the average number of primary factors which enter into each variable. This may be regarded as an advantage ; but in practice it has several disadvantages. First, it increases the number of factors required to explain the data, for the secondary factors have to be added to the primary. Secondly, the effect of using oblique axes is to reduce the factor loadings of the variables on the primary factors.¹ Moreover, the sum of the squares of the loads on a factor may be regarded as probably the best measure of its descriptive or explanatory power when used by itself ; and in general the effect of rotating the axes to find oblique simple structure is to reduce the average value of this sum also. This, of course, is not a necessary consequence of the use of oblique axes, for if a suitable choice is made, both the average sum of the loads and the average sum of the squares of the loads may be considerably increased. And this is probably what happens when oblique axes are chosen naturally to give each factor as much diagnostic weight as possible, without reference to abstract theoretical conditions. But with oblique simple structure the tendency is in the opposite direction ; and both the average sum of the loads and the average sum of the squares of the loads tend to go down. Mathematically this is of little importance ; for the true value of h^2 may easily be recovered. But this cannot be done, unless explicitly or implicitly we introduce the secondary factors which the obliqueness generates, and allow them to present their contribution to the total variance. On the average, under the conditions contemplated, there tends to be a fall in the contribution which the primary factors by themselves make to the variance ; and hence they tend to lose some of the explanatory power which orthogonal structure would allow them to retain. Therefore, if we rely on the primary factors alone—and in practice we often do so—oblique simple structure tends to be too parsimonious ; for some of the explanation has slipped off the axes out of sight into the angles between them.

An example may perhaps be given. In Cattell's interesting second application of the principle to the field of personality as a whole (3), the average value of h^2 for his 13 centroid variables is .542, whereas the sum of the average amount of the variance accounted for by the eleven factors, rotated to find oblique simple structure, is .342 ; and if two other factors are added on the same lines to the eleven admitted, in order to make the number of factors equal to that of the centroid analysis, the figure² would not rise above .375.

If we wish to avoid this weakness, it is not enough to show that h^2 , our original hold on the variance, is still available ; we must actually recover and use it all. That is to say, the secondary factors have also to be brought to bear on the fundamental variables. The secondary factors refer directly to the primary ones, but they go through them to the variables themselves. So that in the long run the variables have a double set of loadings, one directly on the primary factors, and one indirectly, and in a more complicated fashion, on the secondary factors. Looked at in this way, oblique simple structure, even if attainable, does not promise outstanding gains in parsimony ; and the bias of scientific method would seem to be in general rather against the acceptance of it as a descriptive and explanatory scheme. It should be justified in each particular case in which it is used.

III. SIMPLE STRUCTURE AND INVARIANCE

Let us now turn to the second point stated at the outset, viz., the claim that simple structure gives invariance. Invariance in some sense is necessary for objectivity ; without it factorial analysis is merely a computer's private amusement, not a branch of science. Factor loadings are not expected to be invariant from one population to

¹ Illustrations of these points will be found in Burt, *Factors of the Mind*, pp. 358f.

² There appear to be some slips—decimal points misplaced—in the corresponding figures, given for his first application of the principle to the 'personality sphere,' in his book (3, p. 313 et seq.). But the table on which the figures are based is given on p. 88 of his article (1) : using these, a recalculation gives .660 as the average value of h^2 and .411 as the sum of the average contributions made to the variance by the oblique factors.

another, if the latter differs from the former in a way affecting the factorial composition ; but if there is no such difference in populations, the factorial composition of a variable must remain the same, when it is moved to another battery involving the same common factors. Moreover, even when populations differ, invariance may still be secured in a lesser degree. If the same common factors are involved in two batteries, each quality should be explained by the same set in both cases, though the loadings may differ. Or, to put the point in another way, if analysis of one battery shows two variables to have certain factors in common, the analysis of another battery which also contains the two variables should also show them to have exactly the same factors in common and no others.

Now, it is claimed with justice that under certain circumstances simple structure can secure the fulfilment of this condition. If two batteries are found to resolve into a simple structure showing the same common factors, then any variable common to the two batteries will manifest the same factorial composition. But here two important qualifications must be made : (1) the common factors in the two cases must be the *same*, and (2) the invariance claimed and achieved has nothing to do with the simplicity of the structure ; it depends solely on the identity of the factors. If the variables occur in two batteries, and if those factors present in the one battery which affect both variables in question are the same as those which affect them both in the second battery, then the factorial composition of these variables in respect of their common elements must also be the same ; and invariance is secured. If the analysis does not give this result, then either the composition of the variables has changed or the factors alleged to be the same are not so in reality. The main consideration, therefore, is that the common factors in the two cases must be the same ; and it is in the certainty which simple structure is alleged to give in this regard that its main virtue would seem to lie. For its exponents allege that there is no other principle to guide the rotation of the centroid axes, and that apart from it the analysis is indeterminate. We do not agree with this contention. But meanwhile we must recognise the strength of the desire for certainty, and the attraction of a method that rings a bell when the rotating axes have reached the right position. But again there are qualifications : it must be the right bell that rings ; and we must be sure that we hear it. These two points require some separate consideration.

The condition of simple structure has been stated very briefly in recent times : it is satisfied if for every variable there is one zero loading on any one factor, i.e., one zero loading in every row of the factorial matrix. But this condition, though it gives simple structure (by definition), does not necessarily give a unique result : it *may* be satisfied by the rotation of centroid axes into more than one position. That is to say more than one simple structure solution may be possible in the analysis of one and the same battery. Thus the bell may ring in more than one position and the desired certainty is reduced, if not lost. To meet this difficulty the conditions must be strengthened : simple structure must be unique. It would seem therefore, that we must go back to something like the earlier position and demand not only one zero entry in each row of the factorial matrix, but also as many zero entries in each column as there are factors and as many zero cross-products for each pair of columns as there are factors.

With infallible material it is not difficult to discover when these conditions are satisfied ; but with fallible material, such as behaviour ratings of sample populations, the matter is not so simple. The error of the correlation coefficients and consequently of the factor loadings has to be taken into account, since few of the coefficients will actually be zero.

As has been pointed out in connexion with the identification of simple structure as a whole (5), it is dangerous to assume that a desired result has been attained and the data can just be brought into

alignment with it by interpreting all the errors in the direction we wish it to go. The characteristic of random error is to vary in all directions, and the only case in which we have the right to interpret all the errors in our favour is when the only effect they can have is to move the value of a variable away from the position which our hypothesis requires it to take up. The position has frequently been stated and need not be insisted on further here. But an important distinction requires to be drawn. If we frame a hypothesis before coming to our data, and on scrutinizing the data find that the departures from the calculated values are such as chance would produce, we can claim, not indeed that we have proved a hypothesis, but that we have confirmed it, and increased its probability. But on the other hand if we form a hypothesis only after scrutinizing the data we cannot confirm it from the data themselves ; for this further data are required. The search of simple structure, in so far as it affects any particular dimension or axis or factor, is in this latter situation. The search is carried out blindly, mechanically, with no psychological hypothesis in mind, and the enquirer gladly hails whatever he gets. But he has then to remember that if a point is at a distance from a hyperplane equal to its P.E., it is equally probable that it should really be in the hyperplane and that it should be twice as far away from it as it actually is. Thus if twenty variables depart from a possible but previously unsuspected axis or hyperplane by an amount equal in each case to the probable error, the most probable true result is that only half of them belong to it. If more than half are required to belong to the hyperplane by our hypothesis, then our hypothesis, framed from the data, is not only unproved, but is not even plausible.

It is safe to say, therefore, that with fallible material the number of apparent zero entries which simple structure requires should be double the number of zero entries with infallible material. But this is not all : the situation is often complicated in two ways. As has been pointed out above, the use of oblique axes tends in practice to reduce the average factor load, and this increases the ease with which apparent zero loads are located. Further, to make certain of securing all the significant material the analyst sometimes carries his analysis beyond the point where error in the residuals can be expected to be as large as the significant material, and extracts one or more factors which necessarily have relatively low average loads and make only a small contribution to the total variance. This, however, increases the ease with which different zero entries of all kinds may be found, and augments the danger that we may delude ourselves into finding simple structure in places where it is not really present. It will cause the bell occasionally to tinkle when no one is at the door. How much numerical weight should be given to all these considerations in the blind search for simple structure it is impossible to say with any accuracy. But we believe that it is considerable.

Another point has to be mentioned, which adds to the difficulties of assessing the true value of a search for simple structure in fallible material without psychological guidance. It has been urged, rightly we believe, against the practice of leaving the factors unrotated, that the factors initially obtained are relative to the battery, and vary with the introduction of new variables. It is not always recognized, however, that the method adopted to find simple structure may be open to a similar objection. The method used is essentially negative. In a space of n dimensions, it consists in finding a set of n hyperplanes each of $(n-1)$ dimensions, such that each hyperplane contains the required number of variables, and exhausts their communality. These variables have no load on the remaining dimension, and they provide the number of zero entries which the conception of simple structure demands. That is to say, each dimension is determined by a set of variables which do not possess it or enter into it ; and those which do possess it, have merely to accept what is offered to them and make the best of it. Thus a factor abundantly present in some particular field will not be acknowledged, unless a sufficient number of variables, having nothing to do with it, have been incorporated in the battery. Further, by a judicious choice of variables, a hyperplane can be built up in almost any direction. An axis can, as it were, be steered about, just as the first centroid axis can, by a manipulator who takes out a variable here and there, and puts others in, leaving the variables, which are not in the hyperplane in any of its positions, to suffer in silence. The variables which are most concerned do not determine the position of the axis which is introduced to

explain them. Of course, no one intends to treat the data in an unfair fashion, but when, as is the approved procedure, a battery is being prepared to show simple structure, can it be guaranteed that some such arbitrary action is not taken? The point is of more than academic interest.

IV. THE PROBLEM OF IDENTITY OF FACTORS

If all these difficulties have been overcome and dangers avoided, there is still one final task before us. Let it be assumed that we have found unique simple structure for the factors of each of at least two batteries: we have still to decide whether the two simple structures coincide, and reveal the same common factors. We would beg the question, were we to assume that they are identical merely because in each case simple structure is present; and the position would be still less satisfactory were we to make that assumption when only a plausible approximation to simple structure is to be found.

At first sight factors may be identified in more than one way; but in the end everything comes back to the original variables: what may be regarded as the ultimate test depends on the way in which these variables are interpreted. Factors are the same in so far as they enter into the same variables in the same way; and the invariance in the interpretation of the basic variables is the guarantee of the identity of the factors. Thus, simple structure, as such, is not a guarantee of invariance in the interpretation of the variables: on the contrary, invariance in the interpretation of the variables is the condition of identity of factorial structure, whether the structure be simple or not.

It may be contended that we should cling at all costs to the conception of simple structure, because we have no satisfactory alternative. Thus, for example, Cattell says, "we have rejected one of the two major approaches normally approved by scientific method—namely that of inventing a hypothesis about the particular factors expected and attempting to discover a factorization to match it—because in this field almost any hypothesis could be so 'confirmed.' Instead we seek general guiding principles for the mathematical analysis itself which will lead to a unique solution." (3, p. 281.) And Thurstone speaks of the situation as "indeterminate" when simple structure cannot be found (7).

The pessimism evinced by this attitude seems due partly to the negative method of identifying factors (or reference vectors) by hyperplanes of $(n - 1)$ dimensions. In this method every journey is a movement into the dark, and we are guided into the right path only by finding that all the wrong paths have been blocked. Consequently, if the blocked paths are opened, we are indeed lost. Partly too, the despair arises from failure to recognize the resources open to a positive analysis, and Cattell seems a little too sweeping in his statement that in this field almost any hypothesis could be so "confirmed."

Before developing the claims of what may be called a more psychological attitude to factor analysis, as opposed to a purely mathematical one, a further limitation of the use of simple structure may be insisted upon. Granting for the moment, and against what seems the psychological probability, that the field of personality as a whole, with all its traits and factors included, could be adequately described in terms of simple structure, a random selection of traits within that total field is not likely to be analyzable in the same way. Several dimensions are probably lacking in any random sample, and it is a matter of chance, or Providence, if the exact supply of hyperplanes to determine the exact number and direction of reference vectors is present or not. If factors corresponding to an ultimate simple structure are contained in a random sample of traits, they cannot in general be found by the methods at present employed to find simple structure, nor perhaps, we venture to suggest, by hyperplane methods at all. If we wish to identify the factors in a limited portion of the field of personality, some positive method of doing so must be adopted.

There is, of course, one way of avoiding this, which Thurstone seems to favour, viz., that of forbidding the use of factor analysis in every limited part of the field, unless the investigator has played Providence and selected the variables so that the only true simple structure will reveal itself. This difficulty has been realized by Cattell, and to meet it he has furnished us with analyses which attempt to cover the whole field and so to provide, in outline at least, the scheme of the ultimate simple structure. Many points in his able and extremely interesting treatment deserve discussion. But at present only certain preliminary considerations can be mentioned.

(a) The first point is one insisted on by Cattell himself: he is not always sure when the bell has rung. In the analysis of the main material which he uses to find the ultimate pattern of the "source traits" of the "Personality Sphere," he indicates three alternative solutions, one involving twelve factors, and two each involving seven: (3, Chap. IX; cf. 2.) He is attracted by all three, and although he comes down in the end—lightly perhaps—on the side of the twelvefold analysis, it is clear that the claim to certainty which gives the conception of simple structure its first appeal is not fully confirmed in practice.

(b) The second point arises from the comparison of the twelvefold analysis mentioned above with the results of another analysis, this time into eleven factors, of the "Personality Sphere" by means of another set of variables (4). The variables in the latter cover in substance the same field as those in the former, and ultimately there is not much in the one which is not present in some degree in the other; but the individual variables in the one scheme often vary considerably in detail from those in the other.

In certain cases, however, there appears to be identity, the same trait appearing in both analyses; and with respect to these traits, we are in a position to ask how far the factors in the two simple structures are the same. We may base our answer on the degree of invariance found in the common variables. For this purpose we may take the following traits: four of them seem identical.

Trait						First Analysis	Second Analysis
(a)	Self-assertive	v.	Self-submissive	..	..	No. 1	No. 4
(b)	Strong-willed	v.	Indolent	..	..	11	17
(c)	Cultured	v.	Boorish	..	..	12	10
(d)	Ascendant	v.	Retiring	..	..	21	18
Two others may be added with lesser certainty of identity.							
(e)	Co-operative	v.	Obstructive	..	..	14	1
(f)	Hostile	v.	Trustful	..	..	24	11

In the two enquiries three of these variables, viz. (c), (d), and (e), are measured in opposite directions: their factor loads have accordingly been reflected in the figures which follow. In the first analysis twelve variables are set forth, in the second, eleven; but two of these in the first analysis are not found in the second, and the second adds one not in the first. Moreover, another factor in the first analysis is claimed to be only partly identical with its nearest opposite number in the first.

V. A CONSIDERATION OF SOME OF CATTELL'S RESULTS

In comparing Cattell's two analyses, it has to be borne in mind that a skilful and prolonged attempt was made to reduce as many of the loads as possible to insignificance; and also that a larger number of factors were extracted in the centroid matrix than were retained in the simple structure. It seems unwise therefore to determine the agreement or disagreement by reference to the coincidence or otherwise of insignificant loads, for there are bound to be many such coincidences by chance, although it is impossible to say how many. We suggest therefore that consideration should be given solely to the significant loads. There is one difficulty, however. Where a load is significant in both analyses, judgment is easy; but where it is significant in one and not in the other, the difference may not be great, and may not be significant. It seems right, therefore, to indicate the level of significance, which the difference of such coefficients attains, and in general to treat them as doubtful cases of little evidential value either way.

In the table which follows, all the significant coefficients are given. Others are added in brackets to make comparison possible. The number of persons in the first

analysis is 208, and in the second 133 ; consequently correlation coefficients over .180 in the former case, and over .223 in the latter, may be regarded as significant at or below the 1 per cent. level.

One further qualification has to be made. The populations are not the same in the two cases, the second having probably been selected on a basis of culture. In 'culture' (No. 10 in the second analysis, and 12 in the first), the second population probably has a smaller variance than the first, and this will reduce throughout in some degree the correlations of all other variables connected with culture. The rank of the matrix will not be affected, but if an orthogonal structure is demanded in both cases, a complete identification of the factors will be impossible (see 6, Chap. XI). On the other hand, if oblique solutions are accepted, the factors may be identified in the two cases, and, as Thurstone has shown (7, Chap. XIX), if there is a simple factor structure in the wider population, the same structure with the same factors, can be found in the narrower population also. The general effect will be that the factor loads on the culture factor, and on the other factors correlated with it in the oblique structure, will tend to be reduced. The structure, however, should remain the same. That is to say, in the case with which we are concerned, some factor loads in the second analysis which we might expect to be significant, may fail to reach the required level, because of the selection. In dealing with the doubtful cases of agreement, we have to keep this in mind.

TABLE I

Factor		A	B	C	D	E	F	G	H	J	J ¹	K	L	M
(a) I	1	-.33			.24	.39				(-.04)				
II	4	(-.10)			—	.58					-.24			
(b) I	11		.47	.21		(.17)		.33	(-.05)	-.26		.23		
II	17		.22	—		.31		.55	.24		(0)	(.14)		
(c) I	12	(-.04)	.41				(-.10)	.22		(-.15)		.30		—
II	10	-.23	(.20)				.26	.32			.24	-.39		-.22
(d) I	21						.20		.23					
II	18						.34		-.33					
(e) I	14	.59				-.24	.22		.32	(-.03)				
I	1	.49				(-.03)	(-.12)		(.07)		-.47			
(f) I	24	-.41				.25	-.29		-.24				(-.16)	—
II	11	(.01)				(-.05)	(-.08)		(.08)				.22	.20

This table gives seven agreements in the first four variables, viz. : (a) E, (b) B, (b) G, (b) K, (c) B, (c) G, (d) F. There are ten disagreements : (a) D, (a) J, (b) C, (b) H, (b) J, (c) F, (c) J¹, (c) K, (c) M, (d) H. And there are three doubtful cases : (a) A, (b) E, (c) A. In the last two variables we find one agreement, viz. : (e) A ; seven disagreements, viz. : (e) F, (e) J¹, (f) A, (f) E, (f) H, (f) L, (f) M, and three doubtful cases, viz. : (e) E, (e) H, (f) F.

To allow for the effects of selection in the second population, we have to make some alterations in these results. If this selection had not taken place some of the factor loads in the second analysis might have been higher. The following changes may therefore be made ; three from doubtful to agreement, viz. : (a) A, (e) H, (f) F. One from doubtful to disagreement, viz. : (b) E ; and three from disagreement to doubtful, viz. : (f) A, (f) E, (f) H. This gives a total for the first four variables of eight agreements and twelve disagreements ; for all six variables, eleven agreements ; fifteen disagreements and five doubtful cases.

Before coming to any conclusion it may be well to compare Table I, thus modified with a table in which the variables are different but the factors are undoubtedly the same. Table II has been drawn up on this basis. To match Table I, six cases have

been taken from the factorial matrix of the first analysis, in which variables resembling one another in some degree, but not identical, are set side by side. The descriptions given below are taken from the titles in Cattell's list.

- (a) (1) Self-assertive *v.* Self-submissive.
 (7) Wilful, egoistic, predatory *v.* Mild, self-effacing, idealistic.
 (b) (3) Wise, mature, polished *v.* Dependent, silly, incoherent.
 (11) Strong-willed, conscientious *v.* Indolent, incoherent, impulsive.
 (c) (4) Changeable, frivolous *v.* Thoughtful, stoic, reserved.
 (10) Demoralized, autistic *v.* Realistic, facing life.
 (d) (27) Infantile, demanding, self-centred *v.* Emotionally mature, adjusting to frustration.
 (28) Changeable, characterless, unrealistic *v.* Stable integrated character.
 (e) (14) Antisocial, schizoid *v.* Outgoing, idealistic, co-operative.
 (19) Spiteful, tight-fisted, superstitious *v.* Natural, friendly, open.
 (f) (26) Restlessly, sthenically, hypomanically emotional *v.* Calm, self-effacing, patient.
 (5) Neurotic *v.* Not generally neurotic.

There is obviously considerable resemblance between the members of each pair, and each variable represents a cluster of simpler traits. But Cattell makes clear that where there proved, on test, to be extensive overlap, the overlapping clusters were reduced to a single nuclear cluster. The traits therefore, are definitely distinct from one another in spite of their resemblance, and they cannot be treated as identical. Table II is drawn up on the same lines as Table I.

TABLE II

		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
(a)	1	—·33			·24	·39						
	7	—·31			·21	·35						
(b)	3		·41	·24				·35		—·26	—·25	(·09)
	11		·47	·21				·33		(—·06)	—·26	·23
(c)	4		—·43	—·36		(—·12)		—·27			·18	
	10		—·39	—·42		—·24		—·31			·20	
(d)	27	—·24	—·24	—·37	·50			—·27		·28		
	28	(·01)	—·42	—·38	·34			—·27		(·04)		
(e)	14	—·59				·24	—·22		—·32			
	19	—·49				·22	—·27		—·31			
(f)	26	(·03)		—·38	·41			(—·06)		(·15)		
	5	—·27		—·32	·26			—·23		·27		

This table shows 22 cases of definite agreement, two cases of probable disagreement and five doubtful cases. The contrast with Table I is striking. In Table II the factors are known to be the same and the variables to be different; yet under these conditions there are twice as many agreements (less than a seventh of the number of disagreements) and the same number of doubtful cases as are found in Table I, where the variables are the same. It would seem to follow that the two sets of factors in Table I show less identity with one another than do the two admittedly different sets in Table II. And although Cattell has found simple structure, or the appearance of it, in each of the two analyses of Table I, he has not discovered the same factors in the two cases: that is to say, the pursuit of simple structure has not so far yielded the identity and invariance which is a requisite of successful factorial analysis.

VI. SUMMARY

The advocates of 'simple structure' have claimed that it provides a criterion for the rotation of axes, because it is unique, because it gives invariance, and because it is in accordance with the principle of parsimony. These claims are examined in turn ; and it is found that with oblique axes parsimony is not usually secured, and that invariance does not depend on simple structure but on identity of factors. Further, identity requires a more rigorous criterion than is commonly applied ; and, in Cattell's comprehensive analysis of the 'sphere of personality,' identity of factors does not seem to have been secured.

REFERENCES

1. Cattell, R. B. (1945). 'Description of Personality : Principles and Findings in a Factorial Analysis.' *Amer. J. Psychol.*, LVIII, 69-90.
2. Cattell, R. B. (1946). 'Simple Structure in Relation to Some Alternative Factorizations of the Personality Sphere.' *J. Gen. Psychol.*, XXX, 59-73.
3. Cattell, R. B. (1946). *Description and Measurement of Personality*. Harrap.
4. Cattell, R. B. (1947). 'Confirmation and Clarification of Primary Personality Factors,' *Psychometrika*, XII, 197-220.
5. Reyburn, H. A., and Taylor, J. G. (1943). 'On the Interpretation of Common Factors.' *Psychometrika*, VIII, 53-64.
6. Thomson, Godfrey (1946). *The Factorial Analysis of Human Ability*. University of London Press.
7. Thurstone, L. L. (1947). *Multiple Factor Analysis*. University of Chicago Press.

A NOTE ON FACTOR INVARIANCE AND THE IDENTIFICATION OF FACTORS

By RAYMOND B. CATTELL

University of Illinois

I. *The Need for Objective Matching Techniques.* II. *Two Possible Approaches.*
III. *The Method of Coincidence of Marker Variables.* IV. *Calculation of Chances
with a Symmetrical Marker Criterion.* V. *A General Formula for Marker Coincidence
Matching.* VI. *Summary.*

I. THE NEED FOR OBJECTIVE MATCHING TECHNIQUES

In factor analysis, both with ability studies and with more general personality variables, the need constantly arises to decide whether a factor found in one population can be identified with that found in another. Throughout the early exploratory researches a rough method of inspection and impression sufficed; but, with the multiplication of factors among the same variables and the possibilities of exactness resulting from larger samples, there has arisen a need and a possibility of more precise methods.

The present note calls attention to the general problem and seeks to grapple with it by one of the many possible techniques that deserve intensive study. First, however, it is desirable to consider these various approaches quite briefly before concentrating on the one developed here.

If one wishes to be mathematically cantankerous, no two factors can ever be said to be 'the same.' Each has reference to a particular population and conditions of testing. Yet psychologically the two may be the same: they are dimensions of variance, due to individual differences in the same psychological quality. Nevertheless, owing to sample differences, the same psychological dimension is likely to have (i) somewhat different patterns of loading manifestation, (ii) different total variances, and (iii) (if they are oblique factors) different intercorrelations in each.

It is the contention of the writer that a factor should be identified by all three indications, allowance being duly made for sampling and test conditions. Some of these, e.g., the intercorrelation of factors, do not apply to every system; but the present paper will be concerned with centroid or multi-factor techniques which have the most general application.

II. TWO POSSIBLE APPROACHES

No statistical tests will be suggested for goodness of match of *variance* or of *obliqueness* to particular neighbour factors, since we shall here concentrate on the *loading pattern* problem only. Two methods have so far been used.

1. The first is to apply a statistical test to express the similarity found for the *whole loading pattern* of the two factors concerned. Commonly this will take the form of applying correlation to the two series of loadings of the same variables, as has been suggested by Burt (1) and carried out by Fiske (5). (A somewhat similar

device has been used in seeking similarities between observed correlations by Spearman (9) and Tryon (13).) But it may also take the form of applying the chi-square test.

2. We may take a certain arbitrary level of loading, and assign to all variables above this level the status of being 'significantly loaded in the factor.' We can then consider two factors identical if substantially the same group of variables is found highly loaded in each and if the same group is found not loaded.¹

The first method has considerable promise, but is limited by the fact that for batteries of less than, say, 30 tests, the standard error of r is bound to be large. However, factor analysis with batteries of less than 30 tests is not very satisfactory anyway. A large and varied battery is necessary to define a factor.

In both methods, but particularly in the first, difficulties arise from the sampling errors of the test reliabilities. As Saunders (7) has shown, loadings reduced by low reliability can be corrected for attenuation by the usual formula; and this correction seems desirable (especially where variables have different reliabilities) before applying correlation. Similarly, wherever requisite and possible, it appears desirable to correct for selection (cf. Thurstone 11, 12).

When both corrections have been made, an index can be used which will simultaneously test matching for (a) loading pattern and (b) agreement of magnitude of variance. The ordinary correlation coefficient will give perfect agreement between loading patterns that are of the same *shape*, regardless of *level*; but a *coefficient of pattern similarity*, r_p , has been suggested (4) which takes account of both shape and level.

III. THE METHOD OF COINCIDENCE OF MARKER VARIABLES

The second method will be dealt with more fully here, not because it is better, but because it is briefer and adapted to exploratory stages, and the writer wishes to answer criticisms arising from his use of it. The quantitative expression here given is implicit in the general non-quantitative practice that a factor must identify itself both by the variables it loads highly and by the variables it excludes (cf., e.g., Thurstone 12).

In principle the method seeks to determine the *probability* that the marker variables found in the first study will be reproduced, to the extent that they are in fact reproduced in the second study, by chance alone. This concerns the probability, first, of that combination of markers being reproduced for any given factor, and, secondly, of comparable reproduction occurring simultaneously for any number of factors pairs in the two experiments.

In the studies referred to by Mr. Greenall in his criticisms (6) (a) twelve factors were found among 36 variables with the men, and (b) another twelve among the same 36 variables with the women. The question concerns which of the twelve factors found for the women can be "identified" with the factors found for the men. In the rotated factor matrices each had substantial loadings (above .25) in about six out of the 36 variables. Initially I took the three highest markers for a factor in the first study and found that these three turned up again in the first six in the second. I argued: "the probability of three given variables out of the 36 appearing in any group of six by chance is well below the one per cent. level: (${}^{33}C_3$ divided by ${}^{36}C_6$ equals $1/357$.)" Later I modified this criterion in favour of a more symmetrical test, and stated that in general the matching of factors "rested on three or more of the six or seven highest variables being the same in the two factors matched" (3). I am indebted to Mr. Greenall for pointing out that these statements involve some

¹ Neither method is capable of avoiding false matches resulting from 'co-operative factors' (2) (8), i.e., factors which are distinct but have the same pattern of loadings *with the variables of that particular battery*. Such factors, however, are of low variance, since no marker or other variable in them can have a loading exceeding .72. Burt's 'symmetry criterion' indicates a possible approach (1, p. 281; *Brit. J. Ed. Psych.*, IX, 1939, p. 68); but does not quite meet the problems here raised.

ambiguity, since the probability of 1/357, worked out for certain factors, in which the first three in one study occur among the first six in the other, does not apply to the more general matching practice finally followed, of recognizing a match wherever three in the first six of one factor are the same as three in the first six of the other.

It is proposed now to develop more explicitly the reasoning concerning these and other relevant probabilities, which have been perhaps referred to a little too casually in the pages mentioned. There were eleven or twelve factors in one study and thirteen or fourteen in the other. But for simplicity we shall consider both studies terminated at twelve factors, so that each is defined by six markers out of 36 variables. The following questions then arise. (1) If twelve groups of three variables each (hereafter called a 3-set) are taken from 36 variables (with replacement after each set), and if twelve groups of six are also taken from the same (or a similar) list of 36 variables, what is the chance of (a) finding the variables occurring in one particular 3-set among those in one particular 6-set, and (b) finding such a match for each and all of the sets in the first twelve with some one set in the second twelve?

Parenthetically I would explain that this "unsymmetrical" procedure (taking the highest *three* in the first study and seeking for them among the highest *six* in the second) was suggested by the phenomenon of regression due to chance errors. Even if the results of the second experiment were not different from the first through any systematic cause, one would expect not exact reproduction, but some degree of regression. Nevertheless, the more symmetrical criterion described below seems preferable, so the unsymmetrical criterion will only be worked out for a limited set of circumstances.

In spite of the claim of my critic that it is 'irrelevant,' I can only repeat that the probability of matching a given 3-set, as indicated in problem (a) above, is given by the expression stated in (3), namely, ${}^{33}C_3 \div {}^{36}C_6$, i.e., the number of possible 3-sets, when three are withdrawn, divided by the total possible number of 6-sets.

This value, which we may call $p_{3.1}$, $= 1/357 = .0028$. Such a match was obtained in the personality researches with Factor G (2) (3), variables 17, 2, and 6 being the three highest in the first experiment, and among the six highest in the second. Since the second G factor is not the only one in the second series to be tested as a possible match, but is merely one out of a possible twelve, the probability we seek is different from the above. We may call it $p_{3.1} \dots$ to indicate that it is the probability of *at least* one match among twelve when the factors are taken at random. This is equal to

$$p_{3.1} \dots = 1 - (1 - p_{3.1})^{12} = 0.033. \quad (1)$$

Thus factor G, and other factors in like case, are reproduced with better than a 5 per cent. probability level.

If all twelve factors had reached this degree of matching, the result would be one that would be likely to occur by chance with a probability of only

$$p_{3.2} = p_{3.1}^{12} \dots = \{1 - (1 - p_{3.1})^{12}\}^{12} = 1.67 \times 10^{-18}. \quad (2)$$

IV. CALCULATION OF CHANCES WITH A SYMMETRICAL MARKER CRITERION

No such high degree of factor invariance, as is indicated by the above criterion, has yet been claimed in any study; and it seems impractical to expect it. Accordingly, it seems desirable to set up a more attainable and more symmetrical criterion, namely, that factors shall be considered 'matched' or invariant *when the most heavily loaded one-sixth of the variables in one factor has at least 50 per cent. of items in common with one-sixth of the variables that are most heavily loaded in the second factor*. In the present studies, with 36 variables this means that the highest '6-set' in one factor shall have *at least* three variables in common with the 6-set of the matched factor.

As before, we may treat the problem as one in which the individual variables are not 'replaced' (since a factor cannot employ a variable twice), but in which the whole 6-set (and indeed *all* variables) are replaced before a second factor (6-set) is taken out (since each

new factor has loadings in all variables). Actually the appearance of a variable as a marker for one factor somewhat reduces its chances of appearing as a marker for a second factor. But where, as in the experiments discussed, all variables have about equal communality, and the loadings in any one factor are only moderate, the simplest approximation is to consider appearance in one factor as not precluding appearance in another.

Accordingly, the answer to the problem akin to that first raised regarding 3-sets, namely, *what is the likelihood of a given 6-set being matched in three or more variables by any one 6-set taken at random*,¹ is the ratio of the number of possible 6-sets, having at least three in common to the total number of possible 6-sets, i.e.,

$$p_{6.1} = \frac{{}^6C_3 \times {}^{30}C_3 + {}^6C_4 \times {}^{30}C_2 + {}^6C_5 \times {}^{30}C_1 + {}^6C_6}{{}^{36}C_6} = .0451. \quad (3)$$

Such a goodness of matching was obtained for practically all factors in the two male populations (2, 3 and reference); but it is not possible to prove this objectively since the wording of the rating scales was not kept identical, certain 'clarifications' having been made between the two studies.

It is obvious that the probability of attaining such a complete series of twelve matches by chance is very small (see below). Indeed, the situation that we are most likely to meet in practice is that in which some, but not all, of the population of factors in one study are matched, to a standard degree, in the second study. To this point we now turn.

V. A GENERAL FORMULA FOR MARKER COINCIDENCE MATCHING

We assume (a) that the same number and kind of variables is used in both experiments, (b) that the same complete number of factors is extracted in both, and (c) that a "marker" criterion, i.e., level of loading, is chosen so low that the appearance of a marker in one factor does not reduce its chances of appearance in another. (Only the last involves an approximation; and a separate enquiry would be needed to decide how far this causes us to overstate the likelihood of getting our actual match by chance.)

For illustration we may use the case over which the debate arose, that in which factors are 'marked' by six variables out of 36, and a 'match' is considered to exist where three or more variables are common to two sets of six. For other cases it is a simple matter to substitute for the value of p given below whatever other value is appropriate to the standard of matching adopted by the experimenter.

If p is the probability that one random marker set in the first series will be matched by at least one among the twelve in the second (for our case, $p_{6.1} \dots = .4255$, by equation (3) footnote above), then the probability that a certain eight pairs and no more will match will be p^8 . The probability of the remainder being non-matches will be $(1 - p)^4$. Hence the probability of such a situation will be

$$p_{12(8)} = p^8(1 - p)^4 {}^{12}C_8; \quad (4)$$

but the probability of *at least* eight matches will be

$$p_{12(8)} = \sum_{N=8}^{12} p^N (1 - p)^{12-N} {}^{12}C_N. \quad (5)$$

Expressed in general terms, with K factors and at least N matching, we have:

$$p_{K(N \dots)} = \sum_n^K p^n (1 - p)^{K-n} {}^K C_n. \quad (6)$$

If all the factors match, this simplifies to $p_{K(N \dots)} = p^K$, or 0.4255 as given by the criterion in the above experiment. A matching for all twelve factors would have a .0000352 probability of occurring by chance.

¹ The probability that a perfectly random 6-set will match *at least* one of twelve 6-sets is, however, $p_{6.1} \dots = 1 - (1 - p_{6.1})^{12} = .4255$.

It will thus be seen that, as between the men's and women's factorizations, the extent of agreement is 7, 8 or 9 matches, according to the rotation accepted (2) (3). The first and published rotation shows 3 or more variables in common to two 6-sets in *seven* cases, namely, *A, B, F, G, H, K*, and *M*; but with alternate factorizations 3 variables also match for factor *E* and factor *I* or *D*. The probability of exactly seven matches by the above formula is 0.13, of eight 0.06, and of nine 0.02. The probability of *at least* eight is 0.08; and, because of the above assumptions, this must be taken as the upper limit of probability of obtaining the demonstrated agreement by chance. It must be remembered, however, that this represents the similarity obtainable as between male and female populations. There is every reason to maintain that the degree of invariance of personality factorizations at present obtainable exceeds chance by a greater amount than this, when like-sex populations are employed.

VI. SUMMARY

The possibility of using marker coincidence (or any other indicator of invariance of the factor-loading pattern as such) along with considerations *beyond* the actual loading pattern, next suggests itself. One is inclined to take into account the magnitude of the factor variance (mean for all variables), and, if oblique factors are used, their correlation pattern with respect to other factors (2). Both of these, as Thomson (10) and Thurstone (11) have pointed out, are very susceptible to changes in the population. Granted similar populations, however, we should hesitate to consider a factor with large variance in one series as a match for a factor with small variance in another. For example, though Vernon and others have pointed out that different studies give divergent amounts of variance for verbal ability in batteries of scholastic tests, they nowhere find verbal ability usurping the place of the general ability factor as the major determiner of variance (14).

If such considerations as mean contribution to variance, pattern of correlation with other factors, or even density of hyperplane suffice, singly or collectively, to provide a tentative first matching, independent of intrinsic marker pattern, of the series of factors in the two distinct experiments, a far higher degree of certainty of final matching results. It now becomes possible to set up this first independent matching of factors, in a given order in the two series, as a hypothesis to be tested later by the criterion of marker coincidence already discussed. And if, among a series of twelve already thus arranged in corresponding pairs, we then find that in each pair three variables are common to the six markers, the coincidence is one the chance of which is far rarer than anything yet computed. In fact it is the probability of twelve simultaneous events, each with a probability of 0.045 (as in equation 3 above). Thus the final probability equals:

$$p_{6,6} = p_{6,1}^{12} = 7.14 \times 10^{-17}. \quad (7)$$

With the comparatively limited reliability of the variables and the resulting attenuation of factor loadings, it is unlikely that matching of personality factor-patterns will, for some years, pass beyond the region of dispute. Matchings are likely to hover around the 5 per cent. level of probability. Progress would be greatly aided if experimenters used exactly the same definitions of variables and similar populations. Out of hundreds of studies recently surveyed by the present writer (2), these conditions have been met only once or twice; and the above example seems the nearest all-round approach to the required conditions. A repetition of the conditions of identical variables, similar populations, complete factorization, and blind rotation of factors should answer the question of the 'reality' of the primary personality factors well beyond a 1 per cent. doubt.

Meanwhile, as far as personality is concerned, some individual factors are matched to a *P* value of 1 per cent., but the total configuration to a value of only, say, 5 to 8 per cent.

The most constantly recurring factors, surest by the above test and also by the weight of less exactly expressible research similarities, are *A, B, C, E, F, G, H, K, and M*.

As far as methodology is concerned, two general approaches have been suggested above—*r* and χ^2 methods, attending to the whole loading profile, on the one hand, and 'marker coincidence,' on the other. The marker coincidence method described here can be developed further by working out the most sensitive number of marker variables and common elements to include in the matching process, as well as in other ways. The total profile method will probably remain the more accurate alternative; but the marker method provides a quick method of assessing goodness of match. It may even recommend itself more as research progresses, and really highly loaded markers can be put in to represent particular factors, and where the factors are so clean-cut as to have negligible loadings in other variables. The division between markers and non-markers would be so sharp as to make the present method highly practicable. It also has a prospect of sharper discrimination, when, as indicated above, the factors in each series can be arranged in order, so that the probability of matching by chance *in the given order* becomes extremely small.

REFERENCES

1. Burt, Cyril (1940). *Factors of the Mind*. London: University of London Press.
2. Cattell, R. B. (1947). *The Description and Measurement of Personality*. London: Harrap and Company.
3. Cattell, R. B. (1948). The primary personality factors in women compared with those in men. *Brit. J. Psychol., Stat. Sect.*, I, 114-131.
4. Cattell, R. B. (1949). r_p and other coefficients of pattern similarity. *Psychometrika* (in press).
5. Fiske, D. W. (1948). Consistency of the factorial structure of personality from different sources. *Amer. Psychol.*, III, 360.
6. Greenall, P. D. (1948). Two criticisms. *Brit. J. Psychol., Stat. Sect.*, I, 64.
7. Saunders, D. R. (1948). Factor analysis: I. Some effect of chance error. *Psychometrika*, XIII, 251-257.
8. Saunders, D. R. (1949). Factor analysis: II. A note concerning rotations of axes to simple structure. *Educ. Psychol. Meas.* (in press).
9. Spearman, C. S. (1927). *The Abilities of Man*. London: Macmillan.
10. Thomson, G. H. (1939). *The Factorial Analysis of Human Ability*. London: University of London Press.
11. Thurstone, L. L. (1946). *Multiple Factor Analysis*. Chicago: University of Chicago Press.
12. Thurstone, L. L. (1945). The effects of selection in factor analyses. *Psychometrika*, X, 165-198.
13. Tryon, C. (1939). *Cluster Analysis*. Chicago: University of Chicago Press.
14. Vernon, P. E. (1940). The structure of human abilities (unpublished).

THE PROGRESSIVE MATRICES AS APPLIED TO SCHOOL CHILDREN

By GERTRUDE KEIR

Psychological Laboratory, University College, London

I. *Problem.* II. *Tests.* III. *Subjects.* IV. *Reliability and Validity.* V. *Item-analysis :*
(a) *Difficulty of Items ;* (b) *Factor Analysis.* VI. *Summary and Conclusions.*

I. PROBLEM

Aim of Investigation.—During the war the Progressive Matrices test was widely used with adults, and the results reported in various publications ; but little seems known about its value as an intelligence test for children. The object of the following note is to present a preliminary report of the results of applying the test to children of school age. It will be found that the figures obtained raise serious questions in regard to the reliability and efficiency of the test as a whole, and the value and arrangement of the several items. These questions require to be satisfactorily solved before it can be accepted as a serviceable school test ; and for this purpose far more extensive investigations will evidently be needed than anything hitherto published.

Nature and Purpose of the Matrix Test. In recent discussions of the mental processes evoked by the test, somewhat misleading descriptions have been given ; and it is therefore desirable first of all to be clear about its real purpose and nature.¹ The Matrix test was originally intended to provide a test of 'intelligence' (i.e., of innate general cognitive ability) which should not be dependent on verbal facility or on educational or cultural background. In its general nature it may be described as consisting of a number of visual problems, each having the form of a two-way Serial Analogies test.

In the more familiar shape, that of a four-term verbal problem, the Analogies test was first introduced by Burt to measure the highest or 'relational' level of cognitive processes ; and proved to yield what was then an exceptionally large correlation with independent estimates of intelligence. As was stated in his original description, it is a test of moderate complexity, involving "(1) the perception, implicit or explicit, of a relation, and (2) the reconstruction of an analogous one by so-

¹ Raven originally described it as "providing a non-verbal series of tests suitable for measuring intelligence" (3, p. 17). But in his latest article (6, pp. 12-13), after citing (among other references) Burt's *Mental and Scholastic Tests*, he observes that "people who construct tests of general intelligence seldom state clearly what they mean by the word" ; and goes on to say that his progressive matrices test "is not a test of general intelligence, and it is always a mistake to describe it as such." He prefers to regard it as a "test of a person's capacity to form comparisons, reason by analogy, and develop a logical method of thinking, regardless of previously acquired information." But with children at any rate there can be no question that the problems are often solved without "forming comparisons," and without anything that can properly be called "reasoning" ; and they hardly ever demand "a capacity to acquire a new way of logical thinking." Such descriptions appear to be based on *a priori* assumptions rather than on introspective reports of the subjects themselves.

called relative suggestion.”¹ Moreover, the introspections showed that, with this and other complex tests, children of equal intelligence might reach a solution in two ways: some, of an ‘analytic’ type, proceed by explicitly discriminating the required relations, often mentally expressing it in verbal terms; others, of a ‘synthetic’ type, trust rather to “an activity popularly described as intuition, whereby we implicitly comprehend the character of the whole, without explicitly analysing it into its component parts, or distinctly formulating their relations” (*loc. cit.*, p. 104): in other words, they get an impressionistic notion of a pattern or Gestalt that is not quite complete; and the solution is achieved by completing the pattern or ‘closing the gap.’

In the verbal form the ordinary analogies test was found to depend partly on a factor of verbal facility. Consequently, attempts were made to cast it into a non-verbal or perceptual shape (cf. 2, p. 222). Yet even with non-verbal material, the simple four-term analogy still seemed easier for the child who tends to rely on an analytic procedure (often verbalized) rather than on synthetic apperception, and provided problems of only a limited complexity. “The higher the level of intelligence, the more complex the problem required to elicit it” (1, *loc. cit.*, p. 95). Accordingly, in endeavouring to construct a test which should be as easy for the one type of child as for the other, Burt suggested a more extended scheme for the presentation of the test problem, so that its formal structure could be rendered more and more complex, and yet at the same time be impressionistically perceived. In describing this version he writes: “the general principle on which the test is constructed may be represented symbolically as follows:

$$\begin{array}{ccccccc} a_1b_1 & : & a_1b_2 & : & \dots & : & a_1b_m \\ :: & a_2b_1 & : & a_2b_2 & : & \dots & : & a_2b_m \\ & \dots & : & \dots & : & \dots & : & \dots \\ :: & a_nb_1 & : & a_nb_2 & : & \dots & : & X? \end{array}$$

where a_1 and b_2 denote two essential aspects of the test-material—shape, colour, line, position, or the like—changing systematically step by step. With small geometrical figures as a basis, an infinite variety of problems, lending themselves admirably to progressive degrees of elaboration, can thus be compiled.”²

It will be seen that the above scheme is that of a *matrix* of rank one; and its order may be anything from 1×1 to 4×4 or higher. Thus, to take a simple type of case, a_1, a_2, a_3 might represent figures with three sides, four sides, and five sides, respectively; b_1, b_2 , and b_3 might represent the insertion of one dot, two dots, and three dots: the bottom right-hand cell is left blank; and in this case the child evidently has to fill it in with a five-sided figure containing three dots. He could

¹ Burt, C., ‘Experimental Tests of Higher Mental Processes,’ *J. Exp. Ped.*, 1911, I, p. 101. The purpose of the analogies test is itself commonly misunderstood: Raven is by no means the only writer who assumes that it was meant to test ‘reasoning by analogy.’ It was, in fact, originally devised in 1907 as one of a series of tests for ‘intelligence’ defined as an integrative process, i.e., for what Stout termed ‘noetic synthesis’ (*Analytic Psychology*, II, chaps. V and VI). Recent writers on the Matrix test usually attribute analogy-tests to Spearman, or describe them as “suggested by Spearman’s noegenetic principles of eduction of relations and correlates.” Actually the twofold relational principle described in the text, was put forward as an *alternative* to the views of intelligence held by Spearman at that time (viz., that it was identifiable with simple discrimination). Spearman himself ascribes the ‘invention’ of the analogies test to Burt (*Abilities of Man*, 1927, p. 201); and the fact that he eventually abandoned his earlier theory in favour of a relational hypothesis is an added indication of the importance of this approach in intelligence testing. The chief difference between the two writers is that Spearman was interested in the relational *content* of the test-items rather than in their relational *structure* or form. It is still perhaps a question calling for decision whether there are special factors for logical ‘form’ or for special types of ‘reasoning.’ Earlier writers (e.g., Bain and Sully) had treated ‘reasoning by analogy’ as a highly special case of association by the relation of similarity: Stout showed that it might be mediated by *any* relation, and that the essential process might be observed at *any* cognitive level. Raven’s description seems to represent a reversion to the views of Sully and Bain.

² Burt, C., *The Backward Child*, 1937, pp. 55–6. The original four-term analogies test was described as “a problem in proportion instead of in words.” Now in mathematics a ‘matrix of rank one’ may be defined as a two-way series of numbers in proportion. Hence the shortened name of the test. (Raven seems to have misunderstood the origin of the name: in his latest article he writes: “Each problem is the ‘mother’ or ‘source’ of an organized system of thought: hence the name” (6, p. 13).)

Progressive Matrices Applied to School Children

do so by explicit comparison and inference ; but as often as not he will do it by taking in the total pattern almost at a single glance.

It is, of course, not essential that such matrices should have a rank of one ; they may be (to quote Burt's nomenclature) 'isoclinal' as well as 'synclinal,' and 'non-symmetric' as well as 'axi-symmetric.'¹ Thus in the harder forms of the test the analogy may arise, not from the relation between the separate terms in a given row, but from the relation between (say) the last term and a combination of the first two and the difficulty may be still further increased by presenting the problem in the form of a 'deranged' matrix, e.g., permuting the terms according to the rules for Latin or Graeco-Latin squares. In constructing the complete or composite test it is, of course, desirable to *proceed methodically*, e.g., to have the different types of structure explicitly formulated, to see that they are equally balanced in number, and to ensure that, in every case, there is 'one degree of freedom' only.

The matrix form has further advantages. It is almost self-explanatory. In the usual form of the analogies test the most troublesome part is to explain to the small child what has to be done : Whipple, for example, when including 'analogies' in the later edition of his *Manual* (test 34A), says : "I want you to find a fourth word which shall have the same relation to the third thing as the second has to the first." With the amplified material, the child, if he can solve the test at all, commonly guesses what is wanted at once. Moreover, by gumming the pictures or patterns on insets in a formboard, the test can be made applicable to very young and even to mentally defective children. This, indeed, was its original purpose.

Like the verbal analogies, the matrices may be either graded (increasing in difficulty) or ungraded (of approximately equal difficulty). And, as with other self-explanatory graded tests, particularly when given with a time-limit, and no fore-exercise, the graded form really constitutes a test of short-distance learning-capacity. It is often this element of progressive self-instruction, much more than their superficial formal nature, that is chiefly responsible for the high correlation of such tests with intelligence (cf. 2, pp. 276). In the earlier experiments it was found that, by suitably modifying both the content and the formal scheme, some of the patterns could still be made to lend themselves more readily to the analytic or explicit type of approach, and others to the impressionistic or intuitive type of approach. The earliest versions of the test, therefore, included two main series of items—an analytic series and a synthetic—almost equally balanced.

As part of a research for an M.Sc. degree at University College, carried out under the general supervision of Professor Burt, Mr. J. C. Raven undertook to select and standardize suitable material.² A large number of pattern-types were suggested by various contributors, including Dr. Stephenson and later Professor Penrose ; and with considerable ingenuity and industry Mr. Raven eventually produced an attractive version of the test, which was subsequently made available in group form.

In the test as now commonly used, there are five types of pattern. (A) Single 'continuous patterns' ; these are virtually 1×1 matrices, and thus not so much analogy tests as abstract geometrical versions of Binet's 'missing feature' test ; (B) 'figure analogies' (as Raven calls them) ; these all consist of four-term (2×2) tests ; (C) nine-term (3×3) tests involving 'progressively altered patterns' ; (D) mostly 'permutations of patterns' (deranged 3×3 matrices) ; (E) also 3×3 matrices, described by Raven as 'analysis of figures into constituent parts,' but really 'isoclinal matrices' involving the addition or subtraction of patterns or both. Each of the five types is represented by twelve items. A special merit of the whole series, it is claimed, consists in the wide range of difficulty covered. The items within each set, and the five sets themselves, were intended to be graded in order of increasing difficulty. The form used in the following study was the 1938 version, published by Messrs. H. K. Lewis & Son. It was given as a group test without a time limit.

II. TESTS

In addition to the Matrix test version, the children were given the Simplex Junior Intelligence Test, the Mill Hill Vocabulary Test³ (1943 version), the Schonell

¹ For illustrations of these various constructional schemes, see below, footnote 1, p. 149, taken from Burt's *Lab. Notes on Test Construction*, 1931, where this type of test is fully discussed.

² Other former members of the Department have also devised Matrix tests. Stephenson has used a form in which the testee *draws* the requisite completion (a form for which higher reliability is claimed : 8, p. 236) ; Cattell has used the device in his 'Culture-free test' ; N.I.I.P. Group Test 70 and Anstey's domino test involve similar principles. It would be instructive to make a comparative study of all these variants.

³ The forms for this test were kindly supplied by Mr. Raven.

Silent Reading Test (Form B), the Burt Graded Spelling Test and the Schonell Mechanical Arithmetic Test (Form B).

III. SUBJECTS

The subjects consisted of 296 children at a primary and secondary modern school in the East End of London, comprising 168 girls and 128 boys, aged $10\frac{0}{12}$ to $13\frac{11}{12}$ years. The majority were tested during May and June, 1946, and a small additional group during July, 1949. Not all the children, however, took all the tests. Where the total number taking a pair of tests was below 100, the figures have, as a rule, been omitted from the correlation tables (cf. Table II below).¹

IV. RELIABILITY AND VALIDITY

Correlations between First and Second Applications. For persons over the age of 13 Raven states that the 're-test reliability' of his form of the test varies between 0.83 and 0.93, declining regularly with age; for a group of children 13 ± 1 years he gives it as 0.88 (6, p. 14: the group was apparently small). Other writers, however, have considered his figures somewhat optimistic. Thus with adults in the Services Vernon and Parry (8, p. 234) state that the test "was somewhat disappointing," and give as one reason its "rather poor reliability" (apparently 0.79 or less). In an investigation with children started in 1939 (left incomplete owing to the outbreak of war) Miss Horton obtained even lower figures with children. On re-testing after an interval of three weeks a sample of 123 boys and girls (aged $11\frac{0}{12}$ to $13\frac{6}{12}$ years) she obtained a reliability coefficient of only 0.71 for the Matrix test as compared with 0.82 for a group test (Northumberland 1934 Series) applied to the same pupils. As she observes, a figure of less than 0.80 is decidedly disappointing, even for a new test of intelligence with children.² With my own data, using the split-half method, the correlation between odd and even items (duly corrected) was 0.76.

It is of still greater interest to know how far the results obtained will remain constant after a longer period of time. We propose to repeat the tests after an interval of several years. Meanwhile, data are available for a small number of the 11-year-old children in the primary school, who had been tested twice at an interval of two years. The correlations between the two successive trials in the Matrix and the Simplex test are shown in Table I. The average correlation is distinctly lower for the Matrix test than for the Simplex; but with these small numbers the difference is scarcely significant.

TABLE I. CORRELATIONS BETWEEN SUCCESSIVE TRIALS

	Matrix		Simplex	
	Number	Correlation	Number	Correlation
Boys	18	.54	21	.79
Girls	23	.74	22	.73
Average	41	.64	43	.76

¹ I am indebted to Dr. Charlotte Banks for generous assistance in the statistical analysis of the data; and I have further to express my thanks to Mr. Summerfield, Miss Sen, and Miss Sinha for help with the computations.

² This is a lenient requirement. Both Vernon (*Measurement of Abilities*, p. 154) and Guilford (*Psychometric Methods*, p. 147) suggest as the lower limit a reliability of 0.90 at least.

Correlations with Intelligence and Attainments. To procure some light on the validity of the Matrix test, product moment correlations were calculated between the several tests of intelligence and the tests of educational attainments. The coefficients were calculated for each age separately ; and, as no significant differences were found between the different ages, the figures have been averaged for the whole age range. The results are given in Table II. The following conclusions may be drawn.

i. Within the age-groups the correlations between Age and the Matrix test for both boys and girls are not significant. The correlations between Age and the Simplex test, though significant, are negligible.

ii. There is a moderate correlation for both boys and girls between the Matrix and the Simplex test, namely, just under 0.60. Raven reports a correlation of 0.855 with the Terman-Merrill test (6, p. 14). But this was obtained from only 150 children, covering a very wide range in age (6 to 13 years), and age apparently was not eliminated : he explains that he intended the test, "not to differentiate between individuals of approximately the same level, but to study the entire range of intellectual development." Miss Horton found that the test correlated by only 0.62 with Burt's revision of the Terman-Binet scale, the slight influence of age being in this case first eliminated by partial correlation. She explains the low figure by saying that "first, the test appears not to be so efficient with children as was originally hoped ; secondly, it appears to depend in part on a spatial factor, while the Binet tests involve a discernible verbal factor." As the figures in Table II suggest, the Simplex test also appears to contain a definite verbal bias.

iii. For the present groups the correlation between the Matrix and the Vocabulary tests is much lower than for Raven's group (0.57).

iv. The correlations between the Matrix test and the educational tests are decidedly poor. With one exception all are below 0.50.

TABLE II. CORRELATIONS BETWEEN TEST RESULTS

Test	1 Matrix	2 Simplex	3 Vocab.	4 Reading	5 Spelling	6 Arithmetic
1. Matrix Boys	..	.588	.413	.577	.324	.356
Girls	..	.588	.360	.494	.484	.314
2. Simplex Boys	.538	..	.684	.729	.725	.508
Girls	.588	..	.749	.670	.569	.614
3. Vocabulary .. Boys	.413	.684	..	.601	..	..
Girls	.360	.749	..	.667	..	..
4. Reading Boys	.577	.729	.601	..	.466	..
Girls	.494	.670	.667	..	.593	..
5. Spelling Boys	.324	.725	..	.466	..	..
Girls	.484	.569	..	.493	..	..
6. Arithmetic .. Boys	.356	.508	..	..	..	..
Girls	.314	.614	..	..	..	..
7. Age Boys	.115	.226	..	..	..	..
Girls	.152	.212	..	..	..	..

In his *Guide* (1947) Raven argues that "the actual age at which a child's ability to reason by analogy first appears is less important than his subsequent capacity to adopt this more abstract form of thinking once it has begun to mature. A child's subsequent educational progress seems to depend largely upon the ultimate degree to which he is consistently able to use this method of thinking, irrespective of the nature of the work in which he is engaged." Were this true, we should expect a much higher correlation with tests measuring the child's "educational progress." But it seems clear that, so far as

progress in the subjects of the elementary curriculum is concerned, the Simplex test provides a much better indication. For this all the correlations are *above* 0.50. Indeed, if the present results can be trusted, it would be very difficult to accept Raven's statement that teachers using the two tests together, Matrix and Mill Hill vocabulary, will be able to assess an individual's capacity to "succeed in any course he or she may wish to pursue."

V. ITEM-ANALYSIS

(a) *Difficulty of Items.* No examination of the relative difficulty of items in the Progressive Matrices test, as given to children, seems yet to have been published. In view of the increasing use of the test for screening entrants to grammar and secondary modern schools, a complete item-analysis would seem to be urgently needed.

The common practice of simply printing the percentage of children at each age who pass each item (as McNemar, for example, does in his study of the Terman-Merrill test ¹) is scarcely adequate. It indicates the order of difficulty, but does not reveal the success with which the items are uniformly spaced. For this purpose the percentages passing each item at different ages should be converted into a single figure, expressing its degree of difficulty in standard measure, i.e., in terms of the standard deviation as the unit.

The point is of some importance, because Raven has recently drawn a somewhat radical conclusion from the asymmetrical distribution of individual scores obtained when his test is applied to adults: this asymmetry, he says, shows that "either intellectual development in childhood is not uniformly proportional to chronological age or the dispersion of ability at maturity does not conform to a Gaussian distribution": hence we must either "abandon the I.Q. method or cease to use the standard deviation as an accurate method of comparison" (6, p. 17). But these inferences only follow if (i) the items are uniformly spaced, (ii) there are enough difficult items to discriminate between the brighter adults, and (iii) the inter-item correlations are sufficiently high for the item-results to be reasonably consistent (otherwise even bright adults may occasionally fail in the easier items intended chiefly for children).

The problem of item-study first arose in the scaling of the items in the original Binet scale. As with the matrix test, the answers to the Binet tests are also marked right-or-wrong, and the child simply passes or fails in each. The same method can therefore be employed with both tests.

When sufficient figures are available to give a fairly complete ogival or sigmoidal 'response curve' for each item, a satisfactory figure can be obtained by adapting Fechner's well-known psychophysical procedure for calculating thresholds, namely, the so-called 'constant method.'² The most recent form of this procedure, which is also in keeping with current statistical theory, has been introduced by statisticians in other fields of work under the name of 'probit analysis.' With this procedure, in order to avoid plus and minus signs, 5 is added to the figure obtained in standard measure; and the scale-unit is then termed a 'probit,' i.e., 'probability unit.'³ 5.0 probits is thus equivalent to 0.0 S.D., and implies a success-rate of 50 per cent. Such a figure means that the difficulty of the item corresponds with the average level of ability of the group showing that success-rate. To be comparable with Slater's diagrams, in Fig. 1 overleaf, more than 5.0 probits imply a higher rate of success, and therefore an easier test.

For adults Mr. Slater has recently published a diagram showing the relative difficulty of the matrix items. He also calls his unit a 'probit.' Apparently, however, he has not actually carried out a 'probit analysis,' based on the 'sigmoidal response curve.' It would seem that he has simply transformed his percentages to standard measure by using the ordinary table of the normal probability integral, and then added 5 to the result (7, p. 20). For teachers and most psychological readers it would

¹ McNemar, Q., *The Revision of the Stanford-Binet Scale*, 1942, pp. 82 seq.

² This was the procedure used by Burt in calculating the difficulty of the items in the Binet scale (*Mental and Scholastic Tests*, 1921, p. 138). His diagram for the Binet items—Fig. 21; p. 139—may be compared with my own diagrams for the Matrix items. It will be seen that the latter show far greater bunching than even the version of the Binet scale in its first and crudest form.

³ A readily accessible description, with a worked example taken from the Binet Scale, is given in *Mental and Scholastic Tests*, 2nd ed., p. 448.

Progressive Matrices Applied to School Children



FIG. 1.

probably have been simpler to leave the figures in standard measure, since to them the term 'probit' is unfamiliar, and might suggest an entirely novel procedure. Further, it would surely have been better to make the scale run in the opposite direction, so that more than 5.0 probits would represent a test of more than average difficulty, instead of representing an easier test. Slater's diagram shows a very considerable bunching of the items for his adult group, and a very unequal distribution of the items up and down the scale. Although the group is "not distinguished for exceptionally high intelligence," there are as many as $18\frac{1}{2}$ items per probit between 5.0 and 7.0, and only $7\frac{1}{2}$ items between 3.0 and 5.0. Nor does his order of difficulty altogether agree with that put forward by Raven.

The scaling of the tests, obtained from the children in my own group, is shown in Fig. 1. To make the results comparable with those of Mr. Slater I have preserved his unit and his mode of scaling. For both boys and girls there is a marked degree of bunching. As with the adults there are far too many items of medium difficulty¹; and marked gaps occur between the extremely easy items and the medium items and again between the medium items and the very difficult. The order of difficulty is nearly the same for boys and for girls (correlation 0.98); and agrees closely with that obtained by Slater for adults (correlation 0.97). On the other hand, the agreement with Raven's order is not so close, even within the sets or themes.

B 8 (a case of identity) is obviously much too easy for its position; so are B 10, D 4, and D 5 (the last two being quite out of place in the 'Deranged' series): E 7 and 8 are far too hard for their position. It should be noted that, where for practical or other reasons a time limit has to be imposed, it is essential that the test items should be in the correct order of average difficulty;² and, since

¹ As has often been pointed out, this is a common defect in the tests in which the item-difficulty has not been checked. Unless special precautions are taken, a more or less random sample of test-items, like a random sample of persons, tends to yield a normal distribution, with bunched frequencies near the midpoint. To secure a linear scale special precautions have therefore to be taken.

² The apparent difficulty of an item depends not only on the item itself, but also on those preceding it: it may therefore alter if the series is rearranged. An unexpected change in operation (e.g., a change to subtracting components after a series of additions, as in B11 and E4, or to a combination of both, as in D5, or to some other subtler type of modification or symbolization, as in D11 and E12, or, with some children, a change back to simpler types after permutations, etc., as in D4) will greatly slow down the responses and increase the failures.

Raven's test has often been used in that way, it would seem better to drop the separation of 'sets' according to 'themes' (especially as themes change within the sets as it is), and keep the series strictly 'graded' or 'progressive.' It will also be seen that, with these groups, the average difficulty of the D-items is actually less than that of the C-items, contrary to the statement in the *Guide*.

A study of the validity of the separate items is urgently needed, but would demand a wide range of intelligence in the children tested. With the present groups the validity-coefficients, assessed by biserial correlation, range from 0.28 to 0.62. By far the best items are the 'deranged matrices' included in set D: the double process of thought here required seems greatly to enhance the value of such items as tests of intelligence. At this level the biserial correlations for the easy items in the harder sets suggest that they really serve no useful purpose. It is probably this feature, together with the excessive number of medium items, that accounts for the unreliability of the test at the 15-30 score range, noted by Fraser Roberts (5).

(b) *Factor Analysis*. The factorial composition of the test as a whole seems fairly well agreed upon. Spearman, it is true, held that perceptual tests contained no group factor either for content or for form; and this seems to have led many writers to suppose that a test like the Matrix might be accepted, without further proof, as "almost a pure test of *g*." Burt, on the other hand, contends that there are group factors of visual perception even in perceptual tests, though they are seldom so well marked as the verbal factors in the verbal tests.

Comparing parallel series of verbal, numerical, and spatial tests (including both 'four-term' and 'serial' analogies), he writes: "In regard to content, the spatial test nearly always has a lower saturation with general intelligence, but it also shows a lower saturation with its own particular group factor: in regard to form, there seem to be small factors¹ common to both the verbal and the non-verbal types where the relational solution is the same." In her study of the Matrix test, as applied to children, Miss Horton reached a similar conclusion. "The Matrix test includes," she says, "(i) the same *general* factor as the other cognitive tests, (ii) a small group factor, connected with *spatial* perception, and (iii) one or more *formal* factors arising out of mental operations required for solving the problems—a factor which seems definitely amenable to experience and training." On factorizing data from adults in the forces, Banks reports that the Matrix test includes a 'perceptual or spatial' factor² (9, p. 87); and Vernon and Parry also state that, "in addition to *g*, it involves to a small extent the visuo-spatial or *k* factor" (8, p. 234).

Less is known about the factors entering into the separate patterns and groups of patterns. Before a composite test can justifiably claim an acceptable degree of "internal reliability or consistency," it must be shown to possess a relatively large general factor entering into all the items, and the supplementary factors should be neither broad nor large: and this entails a "factor analysis of items."³ At first sight no doubt the 60 items in the Matrix test may seem to constitute a much more homogeneous test than the 60 odd items in the Binet scale. But, as Raven points out, the 1938 version has been constructed so as to include five different 'sets' or 'themes'; and "the five sets provide five opportunities for assessing intellectual capacity" (*Guide*, p. 1).

Accordingly, to determine the self-consistency of the several items, tetrachoric correlations were worked out between all possible pairs of items; and the table of correlations factorized by simple summation. The results are shown in Table III. Items in which any percentage entry falls below 10 per cent. have been omitted. The following conclusions may be provisionally drawn.

¹ This has been questioned by G. M. Smith in a well-known enquiry ('Group Factors in Mental Tests similar in Material or in Structure,' *Arch. Psych.*, No. 156, 1933). He, however, relied on Spearman's tetrad difference criterion: if his correlations are reanalyzed by simple summation, they reveal clear evidence of formal factors similar to those described by Burt.

² With the men she also finds that it apparently has a small saturation with the verbal factor (p. 82). This is perhaps to be explained by the fact that (as introspections show) men of an analytic type often solve the harder and more complex tests (D and E) by working out the underlying rule in verbal terms.

³ Cf. Burt, C., 'The Reliability of Assessments of Pupils,' *Brit. J. Educ. Psych.*, XV, 1945, pp. 80f., esp. sect. on 'Reliability measured by Factor Analysis.'

Progressive Matrices Applied to School Children

(i) The first factor accounts for 37 per cent. of the total variance. This is far less than the amount found by Burt and John in the Binet tests ; but the reduction is probably due more to the influence of chance than to the influence of special aptitudes like those entering into the Binet tests. The saturations differ widely for different items. In the main the low saturations would appear to indicate, not so much that the items are unsuitable in themselves (though this seems true in one or two cases, e.g., C 6, C 8, E 4, and probably B 8, E 7, and E 8, which are here omitted), but rather that they are too easy or too hard to discriminate children at this particular level.

(ii) The next two factors are of nearly equal influence, and account for 7 per cent. and 6 per cent. of the variance respectively. The two together subdivide the series into four distinct groups, corresponding (with a few intelligible exceptions) to Raven's 'sets' or 'themes.' The second bipolar divides the whole series into two main groups, namely, (i) C and D, which are of medium complexity and difficulty (positive signs), and (ii) B and E, which are either exceptionally simple or exceptionally intricate (negative signs).¹ The first bipolar then further subdivides both the positive and the negative sections of the second bipolar. In the main it seems to contrast the D and perhaps the B set with the E and C sets. Now the items in the D set (which consist largely of 'permutations') are solved most readily by an 'analytical' and even a verbalized procedure (e.g., finding an explicit rule); the C and E sets, at any rate with children of this age, are solved most readily by a 'synthetic' or 'intuitive' procedure : and so are B 9 and 10.

TABLE III. FACTOR ANALYSIS OF ITEMS

Matrix Item	I	II	III
B5	·469	—·118	—·215
B7	·657	—·253	—·228
B9	·809	+·104	—·316
B10	·663	+·246	—·160
B11	·665	—·206	—·328
Sum	3·263	—·227	—1·247
C4	·669	+·088	+·210
C5	·786	+·220	+·078
C6	·568	+·526	+·229
C7	·827	—·191	+·269
C8	·631	+·165	+·427
C9	·734	+·126	+·171
Sum	4·215	+·934	+1·384
D6	·752	—·316	—·217
D7	·748	—·387	+·222
D8	·646	—·370	+·308
D9	·637	—·219	+·176
D10	·463	—·318	+·104
D11	·319	—·248	+·061
Sum	3·565	—1·858	+·654
E1	·424	+·093	+·242
E2	·532	+·245	—·137
E3	·443	—·109	—·129
E4	·526	+·182	—·352
E5	·455	+·239	—·296
E6	·429	+·398	—·072
E9	·261	+·103	—·047
Sum	3·070	+1·151	—·791

¹ The two reversals seem easily explained. D 6 is as simple as B 7 or 9 ; E 1 is as simple as D 7 or 8.

The general pattern of the table suggests that, with a larger sample of children, the most appropriate form of analysis would be a group-factor procedure with subdivided factors—roughly yielding two broad group factors for the ‘analytic’ and ‘synthetic’ items and four or five smaller factors for the separate ‘themes.’ Meanwhile, the figures for the factor-variances show that the test as a whole is decidedly more homogeneous than the Binet scale. Nevertheless, homogeneity is not necessarily a recommendation. Unless such a test is supplemented by intelligence tests of other types, it might yield a biased rather than an all-round assessment.

In the main the results of the factor analysis agree with those of Miss Horton in her preliminary experiments. In addition to (i) a moderately large general factor, she obtained (ii) a bipolar factor which, she says, “seems to correspond with Burt’s ‘synthetic’ *versus* ‘analytic’ apperception (Meumann’s ‘diffusive’ *versus* ‘concentrative’ type of attention). It contrasts B and D with C and E (and possibly A).” She points out, however, that the construction of items belonging to the same set is not always homogeneous; “what Burt calls ‘isoclinal’ matrices are found in ‘synclinal’ sets and are then especially apt to require explicit analysis.”¹ (iii) In addition she also finds “a ‘difficulty factor,’ due to the high tetrachoric coefficient obtained when an extremely easy item is correlated with a very hard.”

The factor-analysis, like the item-analysis, strongly suggests that some of the items are misplaced. But this needs to be confirmed by a similar analysis with groups of different mental level and supplemented by further introspective analyses.

VI. SUMMARY AND CONCLUSIONS

1. A preliminary study of the Matrix test has been carried out with approximately 300 children, aged 10 to 14. Raven’s 1938 version has been used as being the best known and most widely employed. The results obtained suggest that the claims made for it should not be accepted until the test has been more carefully examined and possibly modified.

2. With children the reliability of the test appears to lie in the neighbourhood of 0.70, and is thus even lower than its reliability with adults. After a lapse of two years the correlation between first and second testings was 0.64, as compared with 0.76 for the Simplex test when applied to the same group.

3. As judged by its correlation with the other tests the validity of the test is also lower than has been claimed. Its correlation with the Simplex test was only 0.56, and with educational tests between 0.30 and 0.60—much lower than the corresponding correlations furnished by the Simplex test.

4. As regards relative difficulty, the test-items are far less evenly spaced than those of the Binet scale. There appear to be too many items of medium difficulty. The order of relative difficulty varies from that given by Raven, but agrees quite closely with that found for adults.

¹ Burt illustrates his general principles of construction by algebraic formulae for the operations required, and by test-items taken from his numerical tests. The following examples will explain his terminology.

Synclinal			Synclinal			Isoclinal			Isoclinal		
Axi-symmetrical			Non-symmetrical						Latin Square		
1	2	3	1	3	4	1	4	5	10	15	5
2	4	6	2	6	8	2	5	7	15	20	25
3	6	(9)	3	9	(12)	3	9	(12)	35	25	(30)

He supposed that in general ‘isoclinal’ matrices would require more explicit analysis than ‘synclinal,’ and ‘non-symmetric’ than ‘axi-symmetric.’ But this does not always hold good: if the rule is too hard for him to find, the child falls back on an intuitive guess, which, with many of the patterns, is apt to yield a correct answer.

5. A factor analysis of the intercorrelations between the items shows that the test as a whole is more homogeneous than the Binet scale, although its own general factor contributes less to its total variance. There are indications of supplementary factors corresponding to the relative difficulty of the items, to the matrix-structure of the different items and sets, and to the mode of solving the problems, which may be either analytic and explicit or synthetic and intuitive.

REFERENCES

1. Burt, C. (1911). 'Experimental tests of higher mental processes.' *J. Exp. Ped.*, I, 93-112.
2. Burt, C. (1921). *Mental and Scholastic Tests*. London : P. S. King.
3. Raven, J. C. (1939). 'The R.E.C.I. series of perceptual tests : an experimental survey.' *Brit. J. Med. Psych.*, XVIII, 16-34.
4. Raven, J. C. (1941). 'Standardization of progressive matrices (1938).' *Brit. J. Med. Psych.*, XIX, 137-150.
5. Roberts, J. A. F. (1944). 'Observations on the efficiency of the matrices test.' (Unpublished.)
6. Raven, J. C. (1948). 'The comparative assessment of intellectual ability.' *Brit. J. Psych., Gen. Sect.*, XXXIX, 12-19.
7. Slater, P. (1948). 'Comment on "The comparative assessment of intellectual ability".' *Brit. J. Psych. : Gen. Sect.*, XXXIX, 20-21.
8. Vernon, P. E., and Parry, J. B. (1949). *Personnel Selection in the British Forces*. London : University of London Press.
9. Banks, Charlotte (1949). 'Factor analysis of assessments for army recruits.' *Brit. J. Psych. : Stat. Sect.*, II, 76-89.

THE TWO-FACTOR THEORY

By CYRIL BURT

Department of Psychology, University College, London

I. *The Origins of the Theory.* II. *The Views and Methods of the Galton-Pearson School.* III. *The Columbia Investigations : Inter- and Cross-Correlations between (a) Mental Tests and (b) Academic Attainments.* IV. *Spearman's First Investigations : the Identification of General Intelligence with General Discrimination.* V. *The Hierarchy of the Intelligences.* VI. *The Analysis of Complete Correlation Tables into Factors.* VII. *The Theory of Two Factors as a Simplification of the Theory of Multiple Factors.* VIII. *Summary and Conclusions.*

I. THE ORIGINS OF THE THEORY

Present Status. In the last number of this *Journal* (37), I sought to show how the factorial methods of the psychologist differ from the procedure first outlined by Karl Pearson, and attempted to defend the various changes introduced. In the present article my object will be to explain how these various modes of approach (more particularly those for which I and my fellow workers have been responsible) differ from that proposed by Professor Spearman. The discussion, I hope, will enable me to reply more adequately to certain criticisms that have recently been raised, and at the same time to clarify some of the obscurities and misconceptions that still hinder a proper understanding of Spearman's own contributions.

In a recent review of Spearman's work Prof. Thomson has rightly complained (34, p. 376) that, although nearly all current research is based on a multifactor theory, "it still remains a fixed belief among the less well informed that every test contains two and only two factors." Certainly, no other hypothesis possesses the same attractive simplicity as the two-factor theory; and, perhaps largely for that reason, most textbooks on educational psychology continue to expound it as the accepted doctrine among British psychologists. "Prof. C. Spearman's view," we are told, "is probably the most suggestive for the teacher." "The results of Spearman's researches have shown that a person's total efficiency is dependent on the relative strength of his *g* and the several *s*'s: *g* is innate, and, unlike the *s*'s, unaffected by practice or training." "The general conclusion to be drawn from all these studies is that the Two-Factor Theory now rests on a firm mathematical basis." Even in the 1941 edition of his well-known manual Sir Percy Nunn declared that, though "Spearman's arguments have not passed without challenge, the Two-Factor Theory holds the field."¹

American writers, on the other hand, and the few Continental psychologists who have touched upon the matter, seem equally agreed that, both as a hypothesis and as a working procedure, the two-factor theory has failed to make good its claims. They point out that "Spearman himself has been obliged to admit, through the postern gate as it were, the third body of factors" (the group factors) "that he had so ruthlessly banished from his fortress." This position is perhaps most trenchantly summed up by Dr. Meili, in what is at the

¹ Cf. Pantou, J. H., *Modern Teaching Practice and Technique*, 1945, p. 56; Mathew, A. V., *Psychology and Principles of Education*, 1939, pp. 188f.; Ross, J. S., *Groundwork of Educational Psychology*, 1944, pp. 235f.; Hamley, H. R., and others, *The Testing of Intelligence*, 1936, pp. 8f.; Nunn, T. P. *Education : Its Data and First Principles* (p. 131) : in his last edition (1945) Nunn discusses at some length the views put forward in my 1917 *Report*, which, as he says, "urged the need for a theory of multiple factors," and, in the light of later work, now seems more disposed to endorse it (see esp. pp. 145-152).

moment the latest discussion *sur la nature des facteurs de l'intelligence* (36). He begins by pointing out that "l'existence de facteurs, autre que le facteur 'g' et les facteurs spécifiques, est aujourd'hui presque généralement admise," and wonders "si on est partout conscient de l'importance considérable de ce fait." He then goes on to speculate why Spearman, together with so many of his followers, "a si longtemps nié l'existence de facteurs de groupe, et a essayé de maintenir sa théorie des deux facteurs, malgré les cas de plus en plus nombreux, où son critère de la tétrade n'était pas satisfait." This is one of the questions we have to answer.

The Non-Statistical Background of the Statistical Psychologist. We shall be in a better position to reply to M. Meili if, before turning to statistical evidence, we first recall the wide difference in general aim and outlook between Spearman and the majority of his opponents. Much of the criticism with which the factor analyst has to contend arises because so many writers, both factorists and their critics, habitually assume that a statistical psychology can be built up on its own foundation, with little or no regard to facts and hypotheses in other branches of the science. No one, I fancy, has suffered more than Spearman from this type of misunderstanding.

It is, for example, generally assumed that Spearman's theory was first suggested to him by certain peculiarities that he had observed, or thought he had observed, in the statistical analysis of mental tests: recent writers have related how, in his first correlational researches, he "noticed" that the coefficients obtained tended to fall into hierarchical order, and how, "seeking the reason for this, Spearman thought it could be explained by the hypothesis that only one common factor was causing the resemblances between the tests" (34, p. 374). His mathematical critics have similarly argued that "Spearman's case for the general factor" is "entirely based on the observation and measurement of hierarchical order among coefficients," and that, "while a general factor is a possible explanation, it is not a necessary explanation of the correlational facts." Consequently, Spearman's proof, it is said, involves a fallacious inversion of a deductive inference (18, 2nd ed., p. 173).

But most of the earlier statistical psychologists looked upon their mathematical calculations merely as aids to verifying, with a greater degree of logical rigour, hypotheses which they had already reached on broader grounds. And Spearman was no exception. The majority, however, were interested, like Galton and Binet, primarily in practical psychology—in the concrete individuals whom they tried to test or study. Spearman, on the other hand, was far more concerned with theoretical problems. And the particular school of theoretical psychology to which he gave a lifelong allegiance was that commonly known as 'evolutionary monism.'¹ Anything that savoured of a belief in multiple factors was for him not merely an error: it was a heresy. His 'proof of the two-factor theory,' therefore, is to be considered, not so much a deductive demonstration, as an inductive corroboration of a hypothesis already considered highly probable on *a priori* grounds.

Alternative Hypotheses. It is indeed scarcely possible to appreciate Spearman's arguments, or even understand his terminology, unless we recall the views and the controversies that were current in his day. He was by no means the first to advocate what he called the 'theorem of intellective unity.' All through the nineteenth century there had been a growing reaction against the pluralistic doctrines of the faculty school; and general psychology had ended by becoming almost aggressively monistic.

In Britain the attack had been led by the associationists. The older empiricists had abandoned all attempts at classifying mental capacities, and had reduced psychological phenomena to a miscellaneous set of 'ideas,' connected solely by accidental 'associations' resulting from individual experience. This Spearman calls the 'anarchic' or 'non-focal' view of mental structure (19, p. 53):

¹ It had a strong following in Germany when Spearman worked there as a student; but in this country it was, as Myers once said, "by that time already a bit out of date." See Hauser, K., *Haeckel und seiner Bedeutung für den Geisteskampf der Gegenwart* (1920).

by way of illustration, he cites some of Thorndike's earlier utterances (8, 9) as apparently supporting it; but it was in fact already obsolete. Influenced by the evolutionary doctrines of the nineteenth century, later members of the associationist school laid chief stress on the continuity of growth and the unitary nature of the organism: mental capacities were depicted as developing, alike in the race and in the individual, by an uninterrupted process of *differentiation* and *integration*. Of this view the leading exponents were Spencer in this country and Haeckel on the Continent. Spencer held that all 'cognitions' had arisen by gradual differentiation out of a single basic activity which he termed 'intelligence'; according to their complexity they could conveniently be arranged in four main levels or classes, but one class merged into another by insensible gradations.¹ Intelligence itself remained the same throughout. Thus, as Guilford has observed, "the conception of intelligence as a unitary entity was a gift to psychology from biology through the instrumentality of Herbert Spencer" (31, p. 459).

At a later date there was considerable dispute as to which of the two main aspects was the more essential. Associationists, like Bain and Sully, were 'analysts,' and regarded differentiation as primary. "The first and most fundamental property of Thought or Intelligence," wrote Bain, "is the Consciousness of Difference or Discrimination." Similarly, Sully declared that "the discernment of difference is the most fundamental and constant element in all intellect: it is known as Discrimination."² On the other hand, the younger critics of the associationist school, like Stout and his followers, laid chief emphasis on 'integration,' and claimed 'noetic synthesis' as the real distinguishing feature of all intellectual processes.

On the Continent the monistic view of mind found its keenest champions among the neo-Kantian school.³ Kant's epistemological doctrine of the 'synthetical unity of apperception' was reinterpreted in a psychological sense. Among German psychologists Wundt was the teacher who exercised the greatest influence on Spearman, although, as we shall see, Spearman's first paper opens with a criticism of the Leipzig school for its preoccupation with abstract processes, remote from everyday life. Wundt himself had dismissed 'intelligence' as a *Sammelname*—a popular notion too crude and ambiguous to be regarded as a scientific concept. Physiological psychologists, he says, had postulated an 'Organ der Intelligenz,' and localized it in the frontal regions of the brain. But 'Intelligenz' is not an elementary activity, like those whose localization has been experimentally demonstrated. We ought rather to consider which among all the various *Komponente dieses zusammengesetzten Prozesses* is the essential *Elementarbegriff* or *Zentralfaktor* that distinguishes activities popularly labelled with the name. The *Zentralfaktor*, he contends, is to be found in the process he calls Apperception; and this, he believes, is a function for which there is a localized cerebral centre—*das Apperceptionszentrum*.⁴

The Hierarchical Structure of the Mind. Meanwhile, among younger psychologists, there were several who, like McDougall, held that zeal for mental unity had been carried to an extreme. Many years before, Lotze had declared: "Man behauptet mit Unrecht, dass die Vielheit der Vermögen der Einheit der Seele widerspreche."⁵ And it was McDougall's aim to develop a more eclectic theory of mental structure by combining what was sound in both the pluralistic and the monistic doctrines. This he thought could be done by developing the notion of the mind as a 'hierarchy of levels, the result alike of the evolution of the race and the development of the individual.'

¹ *Essays*, I, 1868, vii, pp. 321f.; *Principles of Psychology*, 2nd ed., 1872, p. 388.

² *Senses and Intellect*, 1864, p. 325. Sully, J., *The Human Mind*, 1891, I, pp. 61–2. Sully also recognizes 'factors of Integration' (p. 169).

³ In this country Ward was its dominant representative. "Synthesis," he writes, "as Kant was the first to see, is 'the indispensable condition of all experience.' Its recognition has proved to be the revolution of psychology . . . And attention is essential for synthesis." Thus, "instead of a congeries of faculties, we shall assume a single subjective activity, and may call this Attention." The apparent differences in conscious processes are due to differences in the content attended to, not in the activity itself. Hence it becomes possible to "show that all the other faculties are resolvable into attention to as many classes or relations of the objects as are successively presented." Ward, J., *Psychological Principles*, 1918, pp. 60, 66, 69; cf. *Enc. Brit.*, 9th ed., 1886, s.v., 'Psychology.'

⁴ 7, III, pp. 380f.; cf. I, p. 315. On these grounds Spearman himself dropped the word 'intelligence' (26).

⁵ *Medizinische Psychologie*, 1852, p. 150. On most matters of general psychology McDougall considered Lotze a far safer guide than Wundt. McDougall's defence of the theory of multiple faculties is most clearly set forth in *Psychology: the Study of Behaviour*, 1912, pp. 81f.

The term and the idea had been freely used by Spencer. But, whereas Spencer had insisted on his 'principle of unbroken continuity,' so that in his hierarchy the 'stages' or 'levels' merged the one into the other by imperceptible transitions, McDougall, adopting what was later known as the principle of 'emergence,' maintained that the mental processes characteristic of the successive levels differed not only by increasing complexity, but also in quality or kind.¹

Sully, Spearman's predecessor at University College, London, inclined, particularly in his later writings, towards a somewhat similar view. He proposes to substitute the notion of abstract 'mental factors' for that of concrete 'mental faculties'; and also recognizes distinct and different 'levels of mental life.' In his *Textbook* (for most of us at that date *the* British textbook) he begins by urging a more systematic "analysis of the mind into its constituent elements or *factors*"; and, what was then a relatively novel suggestion, goes on to stress the importance of supplementing the old type of subjective analysis by "objective methods." Referring to Galton's recent work, he adds: "the latest indication of this tendency is the introduction of *statistical* enquiry into the science"; and strongly advocates its use. Sully's final chapter has a suggestive section on the 'Measurement of Individual Capacity' (including the 'fundamental constituents of intelligence'); and here he again draws attention to Galton's attempt to plan a 'systematic scheme of psychical measurement.' He then discusses at some length the 'causes of individual variation'; and contends that we may regard all such variations as the "product of two factors." For him, however, the two factors are (i) 'innate capability' resulting from hereditary constitution, and (ii) 'functional exercise.' And once more he supports his view by the contributions of the biometric laboratory.²

II. THE VIEWS AND METHODS OF THE GALTON-PEARSON SCHOOL

General and Special Abilities. This survey of rival hypotheses will help us to appreciate the chief questions in the minds of Spearman and his early opponents. Spearman's first article opens with a lament that "psychologists, with scarcely a single exception,³ never seem to have become acquainted with the brilliant work carried on since 1886 by the Galton-Pearson school" (12). His own immediate object, he tells us, will be to advocate a 'correlational psychology' along these lines; and his 'two new formulæ' (rank correlation and correction for attenuation) are introduced by way of correcting or supplementing what he believes are the chief shortcomings in that 'system.' We must therefore glance a little more closely at the "system proposed by Galton and elaborated by Pearson."

Among writers whose chief interest lay more in individual than in general psychology, the pluralistic doctrine of the faculty school still survived, and furnished a useful working terminology. The version adopted, however, was to some extent modified and moulded by ideas borrowed from contemporary associationists and evolutionists. And it is in the writings of Galton that this more eclectic scheme finds its clearest and most influential statement. In his studies of innate or hereditary

¹ Cf. McDougall, W., 'Physiological Factors in the Attention Process,' *Mind*, N.S., XI, 1902, pp. 329f.; also *Modern Materialism and Emergent Evolution*, 1929, ch.V.

² 3, I, pp. 24, 27, 61f., II, pp. 302f. Sully, it may be observed, holds that these two 'factors' are *not* to be treated like "two mechanical forces combining to produce a resultant" (in accordance with the parallelogram law). In an appendix on 'Classification' he traces back to Plato "the scheme of mental powers arranged in a hierarchy" (p. 327). It was largely owing to Sully's support that Galton's laboratory was reopened at University College a few years later with Karl Pearson as director.

³ (12, p. 96.) Spearman is doubtless thinking here of his fellow psychologists in Germany. In America, Cattell, Thorndike, and their associates had already published correlational studies, and Thorndike's *Mental and Social Measurements* (1904) was virtually an introduction to the work of the Galton-Pearson school. In England, Sully and McDougall, in France Binet and his colleagues, were well acquainted with its more important methods and results.

qualities his chief innovation was the distinction between what he called 'general ability' (or 'general intellectual power') and 'special aptitudes' (or 'special powers').

Instead of attempting to classify individuals, as the phrenologists and faculty-psychologists had done, into an assortment of heterogeneous types, he proposed to start by arranging them on a single linear scale, extending from the dullest idiot to the most brilliant genius. He did not deny the influence of 'special aptitudes'; but he believed that "too much stress is laid on specialties": "numerous instances will show in how small a degree eminence is due to *special powers*." To explain the distinction he cites an instructive analogy. The popular view, he says, is like supposing "that because a youth has fallen desperately in love with a brunette, he could not possibly have fallen in love with a blonde. He may or may not have more natural liking for the former, but it is as probable as not that the affair was wholly or mainly due to a *general amorousness*. It is just the same with intellectual pursuits. Men who have no natural taste for science, and yet succeed in it, may be accredited with sufficient *general ability* to leave their mark on whatever subject it becomes their business to undertake."¹ Often he uses the phrase 'natural ability' as synonymous; and at times he even speaks as though 'general ability' was due almost exclusively to innate or hereditary constitution, and 'special abilities' to interest and knowledge acquired from environmental circumstances.

As Spearman points out at the beginning of his historical review, "the first hints" towards testing for this general ability "came from that suggestive writer, Francis Galton"; and he notes how the types of test that Galton himself favoured were tests of sensory discrimination, such as might be applied under laboratory conditions and furnish precise and objective measurements. That, of course, was in keeping with the views of Bain and most British psychologists at that date. Galton, however, had also devised tests for special aptitudes, though he considered these limited capacities to be of only subsidiary importance as causes of individual variations. Pearson's list of traits is of a more popular character; but in Pearson's view, wherever psychological characteristics are concerned, observational assessments were far more trustworthy than mental tests.²

Correlation and the Study of Causes. The method of correlation was introduced by Galton as an aid both to classification and to the study of causes. Spearman repeatedly reminds us how Galton had claimed that his coefficient would indirectly afford "a measure of the hidden underlying *cause* of the variations."³ Statisticians, however, were not slow to sound a caution. As Fisher points out, in itself a correlation "merely tells us the degree of resemblance." If on independent grounds we are able to specify the cause of such resemblances, then, but only then, the correlation coefficient may be used to assess "the relative importance of the factors which act alike, compared to the totality of factors at work."

Galton's own methods of analysis in many ways anticipated what would nowadays be called an analysis of variance. He had himself observed that, if we try to express the relation between a single observable variable and its assumed hypothetical common cause in terms of such a coefficient (r_{xg} say), then the relative strength of the cause is proportional, not to the estimated correlation, but to its square, i.e., in his notation it is proportional to $\frac{c^2}{c^2 + b^2}$, where (to use modern terminology) c^2 denotes the 'variance' of the 'common factor' (r_{xg}^2), and b^2 the 'variance' of the 'specific factor' (r_{xs}^2). The influence of the residual causes will thus be

¹ *Hereditary Genius*, 1869, pp. 24f. *English Men of Science: their Nature and Nurture*, 1874, p. 126. Galton's criticisms, I take it, are directed against the views of G. H. Lewes, who had attacked Dr. Johnson's 'surprising fallacy' that, in view of their high intelligence, Newton could have written *Othello* and Shakespeare the *Principia* (*Problems of Life and Mind*, 1879, 1st ser., chap. V; *Principles of Success in Literature*, Scott Library, pp. 49f.)

² *Biometrika*, III, 1904, pp. 131-190. This is described by Brown as "undoubtedly the most important contribution hitherto made to the subject of the relation of intelligence to other mental and physical characters" (18, ch. II, 'Historical,' p. 94). It concludes a brief but severe counter criticism of Spearman and his formulæ. Spearman's own discussion of Pearson's work had been based on the earlier Huxley lecture (1903) and the *Grammar of Science* (5).

³ 12, p. 74. The reference is to the opening paragraph of Galton's celebrated paper on *Correlations and their Measurement* (*Proc. Roy. Soc.*, XLV, 1888, p. 135).

proportional to $\frac{b^2}{c^2 + b^2}$. In these two complementary expressions of Galton's we have the germ of the 'two-factor theory' in its statistical formulation.¹

Levels of Correlation. From the foregoing review it will be seen that any psychologist who sought to take Sully's advice, and develop "the new idea of mental tests as a means of gaining more objective information," was confronted with two alternative standpoints. Individual psychology, as represented by Galton in this country and Binet in France, still clung to something very like the old doctrine of multiple faculties. On the other hand, theoretical psychology, at any rate as expounded by what Spearman considered the orthodox school, seemed to have effectively substituted the doctrine of "one unitary cognitive process" only. The next problem we have to consider is therefore this: what difference, if any, would such opposite assumptions make to the correlations we might expect to find?

As Galton pointed out, if two measurable characteristics *a* and *b* are affected by only a *single* common cause, say *c*, the correlation between them will be the product of their respective correlations with that cause, i.e., $r_{ab} = r_{ac}r_{bc}$. And this will hold good when a whole set of measurable characteristics (e.g., arms, legs, body, and head) are affected by the same common cause (e.g., 'conformity to the racial type'). With some of these characteristics, however, certain additional influences may here and there come into play; and any such agency will necessarily augment the correlations between those particular 'organs' to which its effects extend.

These general principles were put in a more explicit form by Karl Pearson. When we examine 'organs' that are 'unlike' (say collar bone and shoulder blade), we find that, in spite of their unlikeness, the measurements for individuals belonging to the same racial type always evince a low but positive correlation (0.10 to 0.50); and when we compare 'organs' that are 'like' (two legs, two fingers, or even a leg and an arm), we find the correlations (0.60 to 0.90) appreciably raised.² Hence, if we draw up a square 'table of double entry,' showing the correlations of each characteristic with every other, we shall find that the rows of figures are (to borrow a term suggested in another connexion by Pearson) distinctly 'heteroclinal.'

In Pearson's view these two generalizations were biological 'laws,' applying equally to mental and to physical characteristics. As we shall see in a moment, Spearman and others held that mental characteristics obeyed different principles from physical. For mental characteristics the consequences of Pearson's twofold generalization are well illustrated by a schematic correlation table set out in the joint article by Spearman and Hart (19). They print *h* for high correlations and *l* for low; and, to illustrate "the results needed to accord with the multi-focal theory, implied in the doctrine of 'levels,' 'faculties,' or 'types' ", they take the case of three 'psychoneural levels': we may call them (i) the sensory level *S*, (ii) the associative level *A*, and (iii) the rational level *R*, and may "suppose that three tests have been selected for each level." Here the faculty doctrine would assume the existence of three special abilities of 'observation,' of 'memory' and of 'reasoning' respectively, each producing high correlations within its own particular group. If in addition there is a 'general

¹ See, for example, 2, pp. 68-70, 114, 223, where the more important extracts from earlier papers are brought together. The main point emerges still more clearly in Pearson's various papers on multiple correlation. As Brown has indicated, Spearman's reference to Galton's formula seems to involve a confusion between the *hypothetical* correlation of one variable with the common factor and the *observed* correlations between two measured variables. (Galton in one place makes the same slip, which in his own copy is corrected in his handwriting.) The student will follow Brown's argument more easily if he compares Guilford's equations (31, p. 365). See also Fisher, *Statistical Methods* (1934), pp. 175-6.

² See, for example, *Grammar of Science*, 1900, pp. 403f. The two generalizations are formulated in terms of 'laws of growth': each individual organism (i) "grows its *unlike* parts more closely associated with each other than with those of another individual"; and (ii) "grows its *like* parts more alike to each other than those of another individual." His views on these 'laws,' however, were more clearly and fully expounded in his earlier lectures at University College, to which my account is considerably indebted.

ability,' affecting both 'like' and 'unlike' traits indifferently, the resulting correlation table would take approximately the following form.¹

TABLE I. HETEROCLINAL CORRELATION TABLE

Test	S_1	S_2	S_3	A_1	A_2	A_3	R_1	R_2	R_3
S_1 ..	—	h	h	l	l	l	l	l	l
S_2 ..	h	—	h	l	l	l	l	l	l
S_3 ..	h	h	—	l	l	l	l	l	l
A_1 ..	l	l	l	—	h	h	l	l	l
A_2 ..	l	l	l	h	—	h	l	l	l
A_3 ..	l	l	l	h	h	—	l	l	l
R_1 ..	l	l	l	l	l	l	—	h	h
R_2 ..	l	l	l	l	l	l	h	—	h
R_3 ..	l	l	l	l	l	l	h	h	—

It was a 'heteroclinal' table of this general type that the earliest investigators, who adopted the methods of the 'Galton-Pearson' school (Cattell, Wissler, and Thorndike, for example), evidently expected to find.² Spearman, as we shall see, was just as firmly convinced that all such inequalities were due, not to 'special abilities' or 'faculties,' but to irrelevant influences, i.e., to what he called 'casual correlation,' which would be smoothed away once the effects of all such disturbances had been removed by appropriate statistical 'corrections.' To understand his argument we must look at one or two actual sets of correlations of the kind that he had in mind.

III. THE COLUMBIA INVESTIGATIONS

INTER- AND CROSS-CORRELATIONS BETWEEN (a) MENTAL TESTS AND (b) ACADEMIC ATTAINMENTS

Methods. The earliest research³ to combine the method of experimental tests with that of correlational analysis was the "momentous investigation by Cattell, and Wissler" (as Spearman has called it). This enquiry merits a somewhat detailed description, because it is so little known. It provided the model on which Spearman's first investigations were framed, and is the one to which he most frequently recurs.

¹ 19, p. 57. In the numerical tables given in my 1909 paper, one of which Spearman reproduces in the same article (p. 54), I had already claimed that there were signs of at least three such clusters of augmented correlations: Spearman's method of disproving this inference we shall come back to in a moment.

² It was to factorize tables of this type that my 'group-factor' method was originally proposed. However, questions about this line of approach I must defer to a later paper.

³ Cattell, J. M., and Farrand, L., 'Physical and Mental Measurements of Students,' *Psych. Rev.*, III, 1896, pp. 618f. Wissler, C., 'The Correlation of Mental and Physical Tests,' *Psych. Rev. Mon.*, Sup., III, 1901, pp. 1-62. Wissler emphasizes that "the conception of the problem" is to be credited to J. McK. Cattell; see also Galton's appendix to Cattell's article in *Mind*, XV, 1890, pp. 373-80.

The Two-Factor Theory

Between 1894 and 1904, following the anthropometric scheme drawn up by Galton, Cattell and his colleagues applied some 22 tests to over 300 students—freshmen at Columbia and women at Barnard College. On the basis of their mutual similarity the tests were classified under four main heads: (i) physical tests; (ii) tests of sensory and perceptual accuracy; (iii) tests of quickness; (iv) tests of memory—visual, auditory, and logical. The higher and more complex mental processes were not included owing to the lack of standardized procedures. In addition, marks were secured for the subjects taken by each student during his college course.

The main interest of the final analysis, as Wissler says in his introduction, lay in the endeavour to discover a method by which “the fundamental elements of *general* and *specific* ability could be isolated and valued.” For the correlational techniques employed, he cites both Galton (2) and Pearson (5). Three main sets of correlations are tabulated and discussed: (i) the intercorrelations between the laboratory tests; (ii) the intercorrelations between the marks for ‘class standing’; and (iii) the cross-correlations between the tests on the one hand and the class marks on the other. His criterion for the presence of a ‘general ability’ is the occurrence of significant positive correlations between the various assessments into which that ability might be supposed to enter. Evidence for ‘special abilities’ was based on a systematic examination of correlations within the four main test-groups mentioned above.

Results. (i) The correlations between the laboratory tests were for the most part decidedly low, but, when significant, almost invariably positive: a few rose as high as 0.38 or 0.39. The largest occurred with pairs of tests that might conceivably involve much the same type of special ability (e.g., visual and auditory memory, quickness in ‘reaction time’ and ‘movement time,’ or accuracy in drawing and bisecting lines). In the main, however, Wissler concludes that the elementary abilities which formed the subject of his standardized tests seem “special and unrelated”; and here, as will be generally agreed, subsequent research has corroborated his view rather than Spearman’s. (ii) The correlations between the tests and the marks for college courses fall, with one slight exception (logical memory), below 0.20. It is accordingly argued that, in spite of the claims of earlier writers, a test which professes to measure memory or accuracy cannot be trusted to measure the abilities thus named for purposes of everyday life; and the discussion leads to a timely warning against what later writers have called the ‘naming fallacy.’ (iii) On the other hand, the correlations between the marks for educational achievements are far higher, averaging 0.56; and Wissler infers that “whatever it is that makes for correlation in class standing seems to hold *generally* for *all* courses.”

TABLE II. CORRELATIONS BETWEEN COLLEGE SUBJECTS

Wissler: Based on 115 to 228 Students

Subject	(i) Latin	(ii) German	(iii) French	(iv) Maths.	(v) Rhet.	(vi) Test	Total
(i) Latin	(.726)	.61	.60	.58	.55	.22	3.286
(ii) German	.61	(.619)	(.59)	.52	(.51)	.19	3.039
(iii) French	.60	(.59)	(.479)	.51	.30	.19	2.669
(iv) Mathematics	.58	.52	.51	(.500)	.51	.11	2.730
(v) Rhetoric	.55	(.51)	.30	.51	(.361)	.09	2.321
(vi) Test	.22	.19	.19	.11	.09	(.049)	.849
Total	3.286	3.039	2.669	2.730	2.321	.849	14.894
Saturation	.852	.787	.692	.707	.601	.221	3.860

$$\text{Divisor} = \sqrt{14.894} = 3.859$$

$$\left. \begin{array}{l} \text{Contribution of} \\ \text{1st Factor to} \\ \text{Variance} \end{array} \right\} = (.852^2 + \dots + .221^2) \div 6 = (.726 + \dots + .049) \div 6 = 45.6 \text{ per cent.}$$

In Table II I give the intercorrelations for the main educational subjects and the test of logical memory.¹ Applying my own method of simple summation, I have calculated the 'saturation coefficients' for the several items, and the contribution of the general factor to the total variance. Spearman, commenting on Wissler's figures, observes that they exhibit only a "limited concordance" with the requirements of his own theory. Nevertheless, as my calculations show, the general factor accounts for about 46 per cent. of the total variance (rather more, namely, 47 per cent., if we exclude the test): this is in close conformity with more recent figures for such subjects. Wissler himself thinks that at least two levels of correlation are discernible, the higher correlations (those between similar subjects, such as 'languages') being attributable to a special ability, and thus 'in accord with expectation.' Incidentally he notes that the values are not unlike those obtained for physical measurements.

To account for the somewhat discouraging results obtained with the experimental tests, Wissler considers several possible explanations. First, "it may be urged that in many cases failure to correlate may be due to want of precision in the results." To meet this difficulty he suggests that "the *precision* of a test may be estimated by *correlating the successive trials*," i.e., by what we should now call a 'reliability coefficient.' The cases in which his own tests were repeated do not, he finds, lend much countenance to this explanation. That, however, is the very point on which, as we shall see, Spearman most strongly disagrees. More important in the eyes of later critics, is the fact that Wissler's data were secured from adult students: students would necessarily form a highly selected group, and consequently exhibit only a very narrow range of intelligence. He reports that "tests similar to those in the Columbia series appear to correlate better with children in elementary schools." The disparity he ascribes to the wide differences in maturity exhibited by growing children—an explanation revived by several later American psychologists. "The brighter child," he says, "will do all things well because intellectually he is more mature," whereas college students are all on the same plane of complete maturity, at least so far as the simpler mental processes are concerned. But the main reason for the lack of correspondence between the laboratory tests and the worldly achievements of his subjects lies, so he thinks, in the limited scope of the tests available at that time. He doubts whether elementary experiments of a laboratory type can really be expected to touch or tap the more complex processes such as "operate on the level of practical life." He therefore pleads for a more "exhaustive canvas of the whole field of human activity, to see if tests may not yet be found that will correlate to a high degree with other lines of activity," e.g., in education and other spheres of daily work. "While the outcome of the research thus tends to negate the *immediate* practical value of such tests [as have so far been used], it suggests the possibility of a solution that will be of great importance"; and, in this urgent field of work, Pearson's correlational techniques should provide "a most promising tool," and "answer questions that can be answered in no other way."

IV. SPEARMAN'S FIRST INVESTIGATIONS

THE IDENTIFICATION OF GENERAL INTELLIGENCE WITH GENERAL DISCRIMINATION

Aims. Unlike that of his predecessors, Spearman's primary aim was, not so much to measure or classify individuals, as to find "a more adequate basis for a unified science." His first contribution is prefaced by a general criticism of Wundt's experimental school for its preoccupation with trivialities (12, pp. 202f.).² So far,

¹ Wissler's table of correlations shows one or two omissions, because students who took certain subjects (e.g., German) apparently did not take certain other subjects (e.g., French or 'Rhetoric,' i.e., Logic and English). I have filled in the gaps by inserting correlations for the missing pairs from London students, adjusting the figures to some small degree to keep the average correlations the same.

² This 'introduction' must nowadays seem a little irrelevant. But both Spearman and Wissler are thinking of Münsterberg's recent tirade (which they both quote) against "the rush to experimental psychology." Wissler had in effect argued that, in addition to the 'search for general laws,' we must also study individual variations: Spearman's plea is for a more adequate 'methodology' in the search for those laws. See also *The Nature of Intelligence and the Principles of Cognition*, especially sections on 'The Present Crisis' and 'The Need of Ultimate Laws' (2nd ed., 1923, pp. 23, 29f.).

he says, it has revealed no "fundamental uniformities or laws," nor yet established any "connection between the psychics of the Laboratory and those of real Life." The "cause of this weakness," he holds, has arisen from looking for 'complete correspondences' instead of for 'partial tendencies.' These can now be measured by Galton's method of correlation. But the "correlational psychology here advocated" (i.e., by Spearman himself) is to be "fundamentally distinguished from the branch now popular as Individual Psychology," since the new proposals will "methodically eliminate individuals as an obstacle to progress, being itself in search of laws and uniformities."

In his search for a "conceptual uniformity" that may supply the "missing link," he proposes to begin with the "cardinal function provisionally termed General Intelligence"—"provisionally," because of Wundt's own criticisms of the term (see above, p. 153). "In the systematic psychology of modern times," so he reminds us in a later volume, such a concept "first attains prominence in the work of Herbert Spencer. Life is taken by Spencer to consist in the 'continuous adjustment of internal relations to external'; and the function or capacity for making such adjustments, so far as these are mental, has been conveniently termed by him 'intelligence'." To have substituted so precise a concept for the nebulous and superficial views of the older faculty psychologists was, so Spearman holds, "a truly surprising achievement."¹

Previous Researches on General Intelligence. The second 'chapter' of his paper is concerned with a "history of previous researches"—an expansion of Wissler's review. He reports as many as twenty-four investigations dealing with the relation between intelligence (or 'general ability,' as the 'Galton-Pearson' school would prefer to call it) and other mental functions in either adults or children.

Of the writers named, the majority had worked mainly with laboratory tests of more elementary processes, such as discrimination or reaction-time—the "German procedure"; the more recent had boldly improvised rough and ready tests of more complex processes, akin to those of everyday life—the "French procedure." The earliest studies, like those of Galton and Gilbert, had announced "a real correspondence between Intelligence and Sensory Discrimination." But many of the later contributors—Griffin, Binet, Stumpf, Ebbinghaus, Schuyten, for example—had thought it more profitable to explore the possibility of assessing 'higher intellectual faculties,' such as attention, apperception, and learning. In nearly every case the sanguine claims of one batch of workers were almost immediately contradicted by the negative results of their successors.

This disappointing outcome Spearman ascribes to "four main faults." Most of the investigators measured neither (i) the actual correlations nor (ii) their probable errors; and those who did, failed to correct their 'raw' correlations both (iii) for irrelevant influences and (iv) for observational error. His own contribution, he says, will be to show how these inevitable disturbances can be overcome by certain new statistical formulæ.

Alternative Hypotheses. He ends his review with a brief summary of the alternative possibilities, so far as they emerge from these results. Here it will be helpful to draw on the fuller statement and the clearer terminology that he puts forward in his later writings.

As his article implied, the fundamental choice, to his mind, lies between assuming either (i) one universal intellective function, "provisionally called general intelligence," or else (ii) nothing but "a number of mental activities perfectly independent": in short, between (i) what he afterwards called the 'unifocal doctrine'—the 'theorem of intellective unity' upheld by most contemporary 'general psychologists,' and (ii) the 'multi-focal doctrine' of the older 'faculty psychologists' and the newer 'individual psychologists,' including Binet in France, Galton and Pearson in England,

¹ (26, pp. 4-6.) The reference is to Spencer's *Principles of Psychology*, 1870, pt. iv. Spearman adds that Spencer's doctrine "was not so much a novelty as a revival . . . Paramount among lay beliefs had been that which assumes all mental ability to lie under the sovereign rule of one great power, namely 'intelligence.'"

and their followers in America, like Cattell, Wissler, and Thorndike. It is noteworthy that he does not explicitly distinguish the third possibility—"the more eclectic view," as McDougall termed it, that embraces *both* the notion of numerous independent abilities *and* the concept of general intelligence.¹

In regard to the *nature* of the 'universal intellective function,' Spearman at that time appears to have thought that only two of the commoner 'sub-theories' deserved consideration in connexion with mental testing: (a) that it is at bottom a simple and elementary cognitive function, best described as 'discrimination'—the view chiefly favoured by Spencer's followers, like Bain and Sully; (b) that it is a more complex, or at any rate a higher function, such as 'attention' (Binet) or 'apperception' (Wundt)—labels, as Wundt and Stout remind us, which designate much the same mental phenomenon (4, 7). The former is the view which Spearman himself unhesitatingly prefers.

But, as his own historical survey indicates, these were by no means the sole rivals. Association or memory, abstraction and the higher thought processes, mental quickness in the literal sense—all had been tentatively put forward as key-processes. And here we perceive a logical flaw in the plan of Spearman's own enquiry. It could never be sufficient merely to demonstrate that the experimental facts were consistent with his own hypothesis, i.e., with the identification of intelligence and discrimination; it was also incumbent on him to prove that they were inconsistent with the remaining possibilities, i.e., with the identification of intelligence with any other mental function.

The Corrected Correlation between Tests and Criteria. Spearman describes his own immediate object as "an inquiry into the exact relation of General Intelligence to Sensory Discrimination, of which we hear so much." In his choice of tests, he says, his "guiding principle" will be "the opposite of that of Binet and Ebbinghaus." The attempt to assess "the more complex mental operations will be unreservedly rejected in favour of the simplest activities," on both theoretical and practical grounds, "as befits the rigour of scientific research." The statistical method he proposes for demonstrating the "correspondence" between the two forms—though it has little in common with factor analysis as we now understand it—constitutes one of his most valuable contributions, and one that has been most frequently overlooked.

Almost all the psychological work of the "Galton-Pearson school" is, so Spearman maintains, hopelessly vitiated by the "neglect to correct for error." No matter how carefully we try to assess such traits as intelligence or discrimination, our assessments can never claim anything like the 'reliability' or 'precision' that Galton and Pearson were accustomed to assume when measuring height or weight. Moreover, the precision (or lack of it) may be different for the two qualities to be compared. With multiple determinations for each, there are two obvious courses—an 'empirical' or a 'theoretical.' (i) We can start by averaging the marks (a) for intelligence and (b) for discrimination, and then correlate the two sets of averages. (ii) Better still, in his view, we can average, not the marks, but the cross-correlations. Neither method by itself, however, can be expected to yield the maximum correlation:

¹ In a letter written in 1913 he explained that such a doctrine "would still be multifocal, for it merely treats Intelligence as a kind of arch-faculty." But ten years later he seemed definitely to be inclining towards an 'eclectic' rather than a 'unifocal' theory (cf. 26, pp. 72f.): indeed, one reviewer suggested that the "change of labels in the book"—'unifocal' to 'monarchic,' 'multifocal' to 'oligarchic,' etc.—really covered "an unacknowledged conversion from a unifocal to a multifocal view."

owing to the 'attenuation' inevitably produced by the 'unreliability' of the data, the values will be far too low. Spearman therefore argues that the figures so obtained should be corrected by means of two new formulæ which he himself proposes, and hopes that the two methods of dealing with the same data will serve to corroborate each other. If both 'corrected' correlations turn out to be unity, then we can conclude that what the *tests* are measuring is identical with the mental quality estimated by the *criterion*.

The idea (as it was put) that "somehow good correlations can be extracted from bad experimental data by sheer mathematical manipulation" called forth strong criticism both from statisticians like Pearson and from experimentalists like Myers (15). Spearman, however, insisted that, in the psychological field, it was "no longer possible to hold up the Galton-Pearson school as a model for imitation." Nevertheless, in the light of later developments, it may be instructive to consider what methods would be employed, were Pearsonian principles adopted for solving the problem that Spearman has here raised. We should, I imagine, proceed by *weighting*¹ the measurements or the correlations, and we should determine the most appropriate weights by the *method of least squares*. Such an approach would lead to a mode of treatment more in harmony with current statistical procedures. Nowadays, I imagine, most investigators would suggest either (i) correlating not the averages but the factor-measurements, or else (ii) computing what is now commonly called the canonical correlation (Pearson's 'bi-multiple' correlation).

Spearman questioned both the legitimacy and the feasibility of any such procedure: the "central function," he declared, "is in no sense an average (14, pp. 61f., and *Psych. Bull.*, XXXVIII, 1941, pp. 818): and, even if it were, "the proposed procedure would violate all the postulates which alone render averaging legitimate." Yet fifteen years later, in the appendix to *The Abilities of Man*, he seems largely to have come round to a somewhat similar method, which he says "is not theoretically inconsistent with weighting according to regression for multiple correlations" (p. xx); he claims, however, that his equations "permit a more illuminative mode of calculation than the regression equations as customarily used in psychology." Nevertheless, the weighting methods which he suggests in the appendix are, I think, never actually used in the text. Indeed, chapter V, taken from a paper written at an earlier date, includes a long attack on any attempt to measure general intelligence by means of an average or sum—including "the elaborate mathematics of standardization, calibration . . . multiple and partial coefficients, etc." (26, pp. 59-66).

Armed with his correctional formulæ, Spearman holds that it is no longer necessary to undertake, as Pearson and Cattell had done, "unwieldy experiments based on hundreds of subjects": henceforth we can be content with "two or three dozen" (12, p. 101). His own research follows the same general lines as that of Wissler, but uses fewer persons and fewer tests. He relies mainly on two groups of children—pupils at a couple of schools near Oxford, where McDougall was at that time Reader in Mental Philosophy.

1. The first consisted of the 24 oldest boys and girls (aged 10.0 to 13.7) in a village school. Ratings for Intelligence were obtained from teachers and others; and three tests of Discrimination were applied—for Weight, Light, and Pitch. (a) The correlation between the average order for Intelligence and the average order for the three tests proves to be only 0.66; but the 'empirical' correction-formula raises this to 1.04. (b) The average intercorrelation between the tests was 0.22 and between the two main ratings for Intelligence 0.58; the cross-correlations between tests and ratings averaged 0.38. The 'theoretical correction', r_{pq} , consists in dividing the mean of the cross-correlations (viz., 0.38) by the mean of the two 'reliabilities' (viz., $\frac{1}{2}(0.58 + 0.22) = 0.40$). This would give for "the true correlation between General Intelligence and General Discrimination" (as Spearman terms it) a value of 0.95. However, when (as here) one of the reliabilities is very low, Spearman prefers to average the 'reliability' by taking the geometrical mean rather than the

¹ The first attempt to use weighted test-measurements in this way was, I believe, that briefly described in *Eug. Rev.*, VI, p. 151: (the figures were taken from an L.C.C. Report). There, in discussing substitutes for the Binet-Simon tests, I argued that the ideal way of procuring assessments for the general factor of intelligence would be to take standardized measurements obtained from 'graded' tests, and calculate the weights by the principles adopted for partial regression coefficients. In actual practice, when constructing booklets of group tests, the number of items was generally arranged so that it would roughly correspond with the proportional weight; and, when using individual tests, the ranks or standardized measurements were simply summed.

arithmetical¹: that generally yields a slightly higher value. In this way Spearman arrives at a figure that is almost exactly 1.00.

2. The second group consisted of 22 boys (aged 9.5 to 13.7) from a 'high-class preparatory school.' Only one test was used (pitch-discrimination, applied 'collectively'). General intelligence was estimated by taking examination marks for "four branches of study." (i) With the 'empirical' procedure the correlation between the two sets of averaged rankings here turns out to be 0.72, and becomes 0.96 on correction. (ii) The 'reliability' of the marks, estimated by averaging the correlations between the four school subjects, comes to 0.71; the 'reliability' of the test is based on its correlation with Music (0.40); and the average of the cross-correlations works out at 0.56. (The reader can check the figures from the detailed inter- and cross-correlations which I have reproduced in Table III.) On applying the 'theoretical' formula, we get a 'corrected' cross-correlation of $r_{pq} = 1.04$; and the average of the 'empirical' and 'theoretical' determinations yields "a final correlation of precisely 1.00."

TABLE III. CORRELATIONS BETWEEN MARKS FOR SCHOOL SUBJECTS
(ADJUSTED)

Spearman : Based on 22 Preparatory School Boys

Subject	1 Classics	2 French	3 English	4 Maths.	5 Test	6 Music	Total
1. Classics	(.920)	.83	.78	.70	.66	.63	4.520
2. French	.83	(.786)	.67	.67	.65	.57	4.176
3. English	.78	.67	(.646)	.64	.54	.51	3.786
4. Mathematics	.70	.67	.64	(.562)	.45	.51	3.532
5. Test (Pitch)	.66	.65	.54	.45	(.446)	.40	3.146
6. Music	.63	.57	.51	.51	.40	(.415)	3.035
Total	4.520	4.176	3.786	3.532	3.146	3.035	22.195
Saturation (Simple)	.959	.887	.803	.750	.668	.644	4.711
Saturation (Weighted)	.958	.882	.803	.750	.673	.646	4.712

$$\text{Divisor} = \sqrt{22.196} = 4.711$$

$$\left. \begin{array}{l} \text{Contribution of} \\ \text{1st Factor to} \\ \text{Variance} \end{array} \right\} = (.959^2 + . . . + .644^2) \div 6 = (.920 + . . . + .415) \div 6 = 62.9 \text{ per cent.}$$

"We thus arrive," says Spearman, "at the remarkable result that the *common and essential element in the Intelligences wholly coincides with the common and essential element in the Sensory Functions*" (his italics); and, in so doing, we have discovered a quick and objective procedure for "diagnosing the *Central Function*" (a phrase of Wundt's), namely, "a few minutes' test with a monochord."

If, however, the 'common element' underlying the marks for the school subjects is *wholly* identical with the 'common element' in the tests of Discrimination, then, it would appear, the influence of any alleged special abilities must be 'vanishingly minute.' Admittedly the correlations between diverse manifestations of intelligence vary considerably; but that is because those manifestations occur at different levels in the intellectual scale. If 'musical ability' does not provide so good an estimate

¹ In its original form Spearman's equation required the arithmetical mean to be taken for both numerator and denominator. This indeed was more consistent with his theory. In the *Am. J. Psych.* articles, however, the arithmetical mean is given for the cross-correlations (in the numerator) and the geometrical mean for the reliabilities (in the denominator): the change was apparently made on the advice of Lipps, who argued that, *streng genommen*, the reliability coefficients should be treated as ratios. In the later *Zeitschrift* article (13) Spearman uses arithmetical means for both throughout; and claims that, at any rate with smaller samples, this procedure is preferable (*vorteilhafter*). The figures cited in the text (0.58 and 0.22) are taken from the first article (p. 91), and seem more accurate than those given in the second article, where the calculation is repeated (p. 269).

as 'mathematical ability,' nor 'mathematical ability' as 'an examination in the classics,' that is not because one depends pre-eminently on a faculty for music and the other on a faculty for mathematics, but merely because (to borrow a metaphor from Galton) the former are not so fully 'saturated' with 'innate capacity' as the latter.¹

V. THE HIERARCHY OF THE INTELLIGENCES

The Disproof of Pearson's Correlational Scheme. This conclusion was fully in keeping with the prevailing doctrine of mental unity, and with Bain's view that Discrimination is the 'fundamental property of Intelligence.' But it was in sharp contradiction with the views assumed by Galton, Pearson, Binet, and the other pioneers in individual psychology whose work Spearman had just discussed. Accordingly, to clinch the argument in favour of an 'intellective unity,' Spearman briefly examines his set of intercorrelations for 'intelligence' as assessed by examination marks (see Table III above), in much the same way as Pearson had examined the available correlations for physical measurements.

In the table for the preparatory school the first three items form a group of 'like' subjects (languages); the last two form another (auditory appreciation); and mathematics, as Spearman reminds us, is commonly supposed to depend on yet another "entirely independent faculty." According to the 'Galton-Pearson school,' therefore, there should be a cluster of exceptionally high correlations for the first group and another for the last. But on turning to the figures nothing of the sort is to be found. Take Classics: its highest correlation (in round figures) is about 0.80, and the rest of its correlations diminish by about 0.05 at each step; there is no sudden drop as we pass from Languages to the Musical abilities. Now take Mathematics (beginning with the largest figure observed, 0.70): the same regular diminution, by about the same amount, is repeated. Or take finally the test for Pitch Discrimination: the figures begin at about 0.65, and still go down by about 0.05, ending with 0.40, as required. As Spearman observes, "the uniformity is nearly perfect," "the unbroken regularity astonishing." Here therefore the trend of the coefficient shows a far closer resemblance to the diminishing figures which Galton deduced to illustrate the 'hierarchy of kinship' than to the pattern of high and low correlations which the doctrine of multiple abilities would require; and an evolutionist might be tempted to extend the principle and to infer that all cognitive abilities are developed "by gradual and growth differentiation out of one primordial function," much as the members of the same family are descended from one common ancestor.²

The Effect of the Corrections. Spearman's table has been so frequently cited³

¹ *Loc. cit. sup.*, pp. 280f.

² McDougall was indeed disposed to accept this latter suggestion. However, Spearman's view of the 'hierarchy of the intelligences' differs from McDougall's, in that McDougall's involved, not a smooth 'merging,' but a more or less discontinuous 'emerging.' In short, as McDougall put it, the two-factor theory is "too simple to interpret the correlational facts" (30, p. 94). My own description of levels within a hierarchical structure was based on McDougall's (14, pp. 165, 169: cf. *J. Exp. Ped.* I, 1912, p. 263). But the divergences between McDougall's view and Spearman's only became clear in the course of later discussions.

Educational writers, noting that Spearman's 'hierarchy' related primarily to educational subjects, seem to have supposed that he had taken the idea from Pearson's revised version of Comte's 'hierarchy of the sciences' (cf. 5, pp. 509-18, where the arrangement of the 'abstract' branches of knowledge is also based on modes of 'discrimination'). Actually both were independently drawing on the ideas of Spencer and his school, which had become common property at the time.

³ It is reprinted by Brown and Thomson, *Essentials of Mental Measurement* (1925, p. 165), by Clark Hull, *Aptitude Testing* (1928, p. 196), and Thouless, R., *General and Social Psychology* (1948, p. 432), and is given, as the authors put it, to illustrate Spearman's "discovery" of the hierarchy and his "method of proving the existence of a general factor."

as an early and typical case in which hierarchical arrangement was actually 'observed,' that it is desirable to make its source quite clear. Most of those who quote it seem unaware that the coefficients given in the table are *not really* 'observed' correlations, but derivative values obtained by applying corrections for age and the like.¹ The correlations actually observed can be readily computed from the mark-lists given in Spearman's appendix. They are set out below in Table IV.

The first factor still accounts for about 60 per cent. of the variance; and now seems to correspond very closely with age. But the diminution of the figures is far less smooth and regular. Moreover, it is manifest at once that the corrections have masked several relevant features. Contrary to Spearman's statement, pitch-discrimination now appears to correlate more highly with music than with any other subject;

TABLE IV. CORRELATIONS BETWEEN MARKS FOR SCHOOL SUBJECTS (UNADJUSTED)

Subject	(i) Classics	(ii) French	(iii) English	(iv) Maths.	(v) Test	(vi) Music	Total
(i) Classics	(.823)	.925	.891	.870	.177	.201	3.887
(ii) French	.925	(.885)	.867	.882	.214	.255	4.028
(iii) English	.891	.867	(.834)	.880	.117	.321	3.910
(iv) Mathematics	.870	.882	.880	(.814)	.196	.221	3.863
(v) Test (Pitch)	.177	.214	.117	.196	(.073)	.381	1.158
(vi) Music	.201	.255	.321	.221	.381	(.123)	1.502
Total	3.887	4.028	3.910	3.863	1.158	1.502	18.348
Saturation	.907	.940	.913	.902	.270	.351	4.283
Age	.898	.811	.810	.780	-.094	.144	

$$\text{Divisor} = \sqrt{18.351} = 4.284$$

Contribution of 1st Factor to Variance = 59.2 per cent. r_{pq} (corrected) = 0.367.

TABLE V. PARTIAL CORRELATIONS BETWEEN MARKS FOR SCHOOL SUBJECTS (AGE ELIMINATED)

Subject	(i) Classics	(ii) French	(iii) English	(iv) Maths.	(v) Test	(vi) Music	Total
(i) Classics	(.687)	.765	.633	.617	.595	.164	3.461
(ii) French	.765	(.704)	.614	.682	.498	.238	3.501
(iii) English	.633	.614	(.585)	.676	.332	.353	3.193
(iv) Mathematics	.617	.682	.676	(.570)	.433	.174	3.152
(v) Test (Pitch)	.595	.498	.332	.433	(.408)	.400	2.666
(vi) Music	.164	.238	.353	.174	.400	(.120)	1.449
Total	3.461	3.501	3.193	3.152	2.666	1.449	17.422
Saturation	.829	.839	.765	.755	.639	.349	4.174

$$\text{Divisor} = \sqrt{17.422} = 4.174$$

Contribution of 1st Factor to Variance = 51.2 per cent. r_{pq} (corrected) = 0.675.

¹ On p. 276 Spearman himself refers to the correlations as 'raw'; but, as the context shows, he merely means that the coefficients have not so far been 'corrected' for unreliability, like those in the last column of his table (p. 291).

but all the observed correlations obtained with the test are now seen to be too small to be statistically significant¹; and r_{pq} (the cross-correlation, corrected for unreliability) is now only 0.367.

Spearman, however, was firmly convinced that irregularities, such as those here discernible, are nearly always due to the two main sources of disturbance that he has emphasized from the outset, first the observational errors that affect all mental measurements, and secondly extraneous influences such as those of differing age. The former, he thinks, are here largely neutralized by averaging the various mark lists; but additional corrections will be required before we can ascertain the "true rank." The latter type of disturbance he attempts to eliminate by "taking the difference between each boy's rank in school and his rank in age"; the result is treated as "a first approximation," and "further corrections" are then applied (*loc. cit.*, pp. 250-1).

However, he evidently felt that this "rather complicated" procedure was not beyond criticism. On a later page he adds: "the whole process could have been avoided; we could have left [the marks] in their raw state, and simply have eliminated the irrelevant factor of Age by the usual formula: indeed, the theoretical precision would have been greater" (p. 267).

Let us then make this adjustment. Using the Pearson-Yule formula given by Spearman, let us partial out the effects of age. The resulting correlations are shown in Table V. The figures, particularly those for Music, are not so high as Spearman's. French, not Classics, now heads the hierarchy; and Pitch and Music still seem to involve an appreciable group factor. There is again nothing like the 'unbroken regularity' and 'uniform descent' which Spearman's table displays. The contribution of the general factor to the total variance is now only 51 per cent.—a value much nearer to what we should nowadays expect; and r_{pq} is still only 0.675.

There can, therefore, be little doubt that Spearman's method of compensating for the effect of age has produced an over-compensation. The device of deducting the *same* figures for age from every mark-list has itself operated as a common or general factor; and the subsequent adjustments appear to have ironed out any minor irregularities that might still have remained. However, the precise effects of his procedure will be seen more clearly, if, instead of trusting (as both Pearson and Spearman had done) to just inspecting the tables as they stand, we apply a formal test of trend.

Meanwhile, let us notice that, in Spearman's treatment of the subject, there is so far nothing approaching what would nowadays be termed factor analysis. Except in referring to external influences, such as Age and the like, the word 'factor' is scarcely mentioned. Spearman's mode of solving his problem—namely, by seeking to determine what is the maximum correlation between a set of tests on the one hand and a set of assessments for the criterion on the other—is much more akin to the calculation of a canonical correlation than to the factorial analysis of correlated measurements into an additive set of uncorrelated components.

VI. THE ANALYSIS OF COMPLETE CORRELATION TABLES INTO FACTORS

Experimental Methods. The ways in which the general method I myself ventured to propose differed from that of Spearman will be most readily understood if we apply it to Spearman's own table. In my view, to establish the presence of a 'general' cognitive ability, and to prove or disprove the existence of 'special' abilities, it was essential to employ tests² for *all* the main levels and types of cognitive process

¹ The original article (12, pp. 291-2) makes it clear that these figures are based on the 22 'musicians only.' In (13, p. 87) they are erroneously described as based on 36 cases, and in (26, p. 139), as "37 boys and girls"!

² The experimental part of the work was carried out at Oxford, and was a continuation of McDougall's experiments on which he had been engaged for some years. Both Spearman's work and my own owed much to his help and suggestions. For the way in which the four main levels in the 'hierarchy' were defined and determined, see the criteria suggested in *J. Exp. Ped.*, I, 1912, p. 263.

(14, p. 199 ; cf, 17, p. 96). This more comprehensive plan made it possible to determine such abilities *by internal rather than by external evidence*, i.e., from the data supplied by the tests themselves, not simply from a comparison with teachers assessments (which, as it proved, generally had a bias towards verbal ability and memory).

Statistical Methods. In deciding whether the correlations between 'like' and 'unlike' characters were 'high' or 'low' Pearson and his followers had relied mainly on impressionistic judgments. 'High' of course must mean *relatively* high, i.e., higher than would be expected ; and, if we are to reach an agreed conclusion about the presence or absence of augmented correlations, we need some systematic procedure for demonstrating that a given coefficient is, in some clearly defined sense, disproportionately 'high' or 'low.' The method I proposed was to seek the 'highest common factor' which would account for all the correlations actually observed, and to fit the observed figures with a reconstructed set of 'theoretical values' deduced from this 'factor' ; the 'deviations' of the 'observed coefficients' from the 'theoretical values' could then be found by subtraction, and their significance formally tested by comparison with the appropriate probable error (cf. 14, Tables V and VI, pp. 162-3, for more detailed illustrations).

The crux of the problem was—how to calculate a set of theoretical values that would accord most closely with the data. It seemed essential that they should be based on the correlation table taken as a whole. The principle eventually adopted was suggested by Pearson's procedure for fitting theoretical values to contingency tables in cases of manifold association.¹ In the contingency table, if t_i is the total of the i th column, t_j of the j th row, and T the grand total for the whole table, the expected value for the i - j th cell will be $\frac{t_i t_j}{T}$; and the significance of the entire set of discrepancies is then tested by the chi-squared procedure.

The 'highest common factor' (the first 'index character,' as Pearson would call it) is thus implicitly treated as a linear function—i.e., as a weighted sum or average—of the several test-measurements, duly standardized. To determine the correlation of each test with this internal criterion, it is unnecessary to compute the factor-measurements explicitly ; we can obtain it direct from the observed correlations by means of the 'simple summation' formula, $r_{ig} = \frac{t_i}{\sqrt{T}}$. The 'theoretical value' of the expected correlation is then equivalent to $r_{ij} = r_{ig} r_{jg}$, in accordance with Galton's product theorem.

By way of illustration the method has been applied to the tables reproduced above.² With Spearman's own correlations (Table III) the first factor alone is statistically significant ; but its contribution to the variance amounts to nearly 63 per cent.—an exceptionally high proportion in the light of later results. The residuals too are amazingly small. Their signi-

¹ Pearson, K., 'On the Theory of Contingency and its Relation to Association and Normal Correlation,' *Biometric Laboratory Publications*, 1904. A more accessible account will be found in Yule, U., *Introduction to Statistics*, 1910, p. 64, eq. (1) : note, for example, that Yule's Table III (p. 66) is in effect obtained by 'fitting a hierarchy' to the observed figures given in Table II (p. 61, 1937 ed., Tables 5.3 and 5.2). As applied to a correlation table the device is equivalent to Pearson's method of 'principal axes' with two important modifications : first, reduced self-correlations are inserted in the leading diagonal ; secondly, simple summation is substituted for weighted.

² For Spearman's data the calculated tables and the factor-saturations are taken (with a few slight corrections) from a paper read to the Liverpool Child Study Society (1913), and later incorporated in my L.C.C. Report on *The Bearing of the Two-Factor Theory on the Organization of Schools and Classes* (1916). I have gratefully to acknowledge permission to use this and other material.

fiance may be tested by the chi-squared procedure.¹ We obtain $\chi^2 = 0.033 \times (22 - 3) = 0.627$. With eight degrees of freedom, this yields P (the probability of getting a value as large as this) = 0.9999. . . . Plainly there can be no question of any supplementary factors. But, as Yule observes, "not only small values of P , but also a value near to unity, should lead us to suspect our assumptions." A fit, therefore, so incredibly close, demands some special explanation. The cause, as I think any impartial critic will allow, is manifestly to be sought in Spearman's method of 'correction.'

He himself evidently realized that the uniformity was exceptional. He notes that the set of correlations obtained at Columbia "manifests only a limited² concordance" with a hierarchical arrangement; and a couple of years later, when reproducing his own table in his joint article with Krueger (13, p. 87), he acknowledges that, in correlation-tables from other schools ("not yet published"), "the hierarchical arrangement of subjects is usually not preserved" ("meistens bewährt sich die hierarchische Anordnung nicht").

Spearman's Criticisms. Spearman took exception to my own method of analysis on two main grounds. First, my notion of a 'hierarchical arrangement' did not conform with his. In an ideal table reconstructed by the product-theorem the 'theoretical correlations' would necessarily be proportional³ throughout. Such a reconstruction he considered to be "highly artificial, if not misleading." Applied to the correlations from the preparatory school, it would not furnish so good a fit as a set of rounded figures "diminishing by equal amounts"⁴; and, when tested by his correction formula, figures so calculated would fail to yield the requisite value of unity, whereas figures "diminishing by *equal amounts*" would give *exactly* 1.00.⁵

These objections will be understood more easily if the reader compares the two reconstructed tables which I have set out below (Tables VIA and VB), which broadly represent the different notions of 'hierarchical arrangement,' from which Spearman and I respectively set out. Table VIA represents a table fitted to Table III in accordance with Spearman's assumptions (equal *differences* between corresponding pairs in successive rows); Table VI is reconstructed from the factor-saturations given at the foot of Table III (based on my own formula, and therefore giving equal *ratios*).

¹ In applying the chi-squared procedure to a table of correlation coefficients I now transform the observed and expected figures to the corresponding z -values ($z = \tanh^{-1} r$): cf. *Factors of the Mind*, p. 339. This accounts for a slight difference between the value here given and that given in my previous discussions.

² Actually, if we judge Wissler's figures (as Spearman judges his own) merely by the absence of reversals, there is only a single anomaly, to which Wissler himself drew attention—namely, "the exceptionally low correlation for French and Rhetoric"—a peculiarity which he thinks is due to an "accidental cause." If we judge by the contribution of the first factor to the total variance (45.6 per cent.), the concordance is certainly more "limited" than Spearman's, but the proportion found agrees far more closely with later results.

³ What Thurstone and others have since called the 'proportionality equation'—"Burt's formula" as Spearman termed it—was first given in my paper of 1909 (14, p. 159); and was in fact merely an algebraic expression of Galton's law of 'proportional dilutions.'

⁴ With the rounded figures, descending by 0.05, the fit is not much better than with my own. However, with those given in Table VI, it is decidedly closer: here, with the isoclinical table (VIA), the sum of the squared discrepancies between the observed coefficients and the expected is only 0.0094, whereas with the synclinal table (VIB) it is 0.0123. But, as later researches amply showed, this result is exceptional, and seems really to prove once again that Table III is highly artificial. It should be added that Tables VIA and VB are taken from a memorandum prepared for the British Association Committee on Factors.

⁵ By figures "diminishing by equal amounts" Spearman presumably means a table in which the rows form an arithmetical progression, diminishing by .05 throughout, as in the table above, which he so often used to illustrate his 1904 results (cf. p. 164; for a more elaborate illustration, see *Brit. J. Psych.*, VIII, 1916, Table IV, p. 175). But it is not necessary for the differences between *different* pairs of columns to be the same; cf. Table VA.

TABLE VI. DIFFERENT TYPES OF HIERARCHICAL ARRANGEMENT
FITTED TO TABLE III

A. Isoclinal							B. Synclinal						
Differences	(i)	(ii)	(iii)	(iv)	(v)	(vi)	Multi-pliers	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	—	·053	·063	·042	·068	·020	(i)	—	·851	·770	·719	·641	·618
(i)	—	·827	·764	·722	·654	·634	(ii)	·851	—	·712	·665	·592	·571
(ii)	·827	—	·711	·669	·601	·581	(iii)	·770	·712	—	·602	·536	·517
(iii)	·764	·711	—	·607	·539	·519	(iv)	·719	·665	·602	—	·501	·483
(iv)	·722	·669	·607	—	·497	·477							
(v)	·654	·601	·539	·497	—	·409	(v)	·641	·592	·536	·501	—	·431
(vi)	·634	·581	·519	·477	·409	—	(vi)	·618	·571	·517	·483	·431	—
Average Correlations							Average Correlations						
(i-iv)	..	..	..	(i-iv)	(v-vi)	·563	(i-iv)	..	..	..	(i-iv)	(v-vi)	·557
(v-vi)	..	..	..	·563	·409	·563	(v-vi)	..	..	..	·557	·431	
Difference ..							Ratio ..						
·154							1·293						
Corrected Cross-Correlations between Reference-Values (i-iv) and Tests (v-vi)							Corrected Cross-Correlations between Reference-Values (i-iv) and Tests (v-vi)						
$\frac{·563}{\frac{1}{2}(.717 + .409)} = 1·000$							$\frac{·557}{\frac{1}{2}(.720 + .431)} = .968$						

Spearman himself was reluctant to be pinned down to any definite quantitative formula. Nevertheless, both his illustrations and his calculations make it clear that, on his assumptions, the rows of a *typical* table would be 'isoclinal' rather than 'synclinal.' The difference can be put quite simply. For the rows to be 'isoclinal,' the *differences* between corresponding figures in any given pair of columns must be constant, that is, with his notation, $r_{ac} - r_{ad} = r_{bc} - r_{bd}$: e.g., taking rows iii, iv, etc., from columns i and ii in Table VIA, we have $.764 - .711 = .722 - .669 = \dots = .053$, as shown in the top line. For the rows to be 'synclinal', the *ratios* between corresponding figures must be constant, i.e.,

$\frac{r_{ac}}{r_{ad}} = \frac{r_{bc}}{r_{bd}}$: e.g., taking the same two columns from Table VIB, we have:

$$\frac{.770}{.712} = \frac{.719}{.665} = \dots = \frac{\text{Sat i } (.959)}{\text{Sat ii } (.887)}.$$

The equality of the differences follows at once from Spearman's formula as he interprets it (eq. 13, p. 85). Writing the means explicitly we have

$$\frac{\frac{1}{2}(r_{ac} + r_{ad} + r_{bc} + r_{bd})}{\frac{1}{2}(r_{ab} + r_{cd})} = 1, \quad (\text{i})$$

$$\text{i.e.,} \quad r_{ac} + r_{ad} + r_{bc} + r_{bd} = 2(r_{ab} + r_{cd}). \quad (\text{ii})$$

$$\text{Similarly,} \quad r_{ab} + r_{ad} + r_{bc} + r_{cd} = 2(r_{ac} + r_{bd}). \quad (\text{iii})$$

Subtracting ii from iii, $r_{ab} - r_{ac} = r_{bd} - r_{cd}$;

Similarly, $r_{ac} - r_{ad} = r_{bc} - r_{bd}$, $r_{ec} - r_{ed} = \dots$ etc.

Q.E.D.

Conversely, a table constructed on this basis will satisfy his criterion with perfect exactitude. On the other hand, a table constructed from the product-formula does not (see last line of Tables VIA and B).

Eventually, however, Spearman conceded that a proportional table might be regarded "as one form [of 'hierarchical arrangement'], though not the only or the most usual form." But as he repeatedly insisted, "in deciding the place of a test or mental process in the hierarchical scheme, we are not concerned with its *actual amount* of correlation, much less with its *proportional amount*, but simply with its *order* or relative rank. . . The problem is not to fit a whole table of figures with one fell swoop, but to discover the orders for the different tests, and show that the orders are the same."¹

In the light of later developments it is interesting to note the continuation of the letter just quoted. He went on to add: "It is true that your proportional equation could, as you say, be deduced from one version of the 'correction formula'" (the reference is to what I may call the Yulean form of equation *f* in the article just quoted, containing geometrical means in both numerator and denominator); "but the equation was never used in that form, as you will see from the article itself." He ends by suggesting that "a better test might be to correlate the *ranks* of each item in the various columns." This suggestion was later amplified into an 'improved criterion'—the intercolumnar correlation between coefficients rather than ranks. By that date the criticisms advanced by Yule and others against the use of arithmetical means in the formula inclined him to look upon the proportional hierarchy as the more typical. But even then (as will be seen from Tables VIA and B) the new criterion would still justify the older, isoclinal type of hierarchy as well as the synclinal, and indeed was expressly designed to do so. The later tetrad-difference criterion, of course, would pass nothing but a matrix of rank one (Table VIB).

Secondly, he criticized no less strongly the "idea of extracting the correlation for one of the tests with General Intelligence by using those very tests themselves as evidence"; and, like Pearson, he particularly objected to any formula which

¹ Letter to the author of 17 March, 1909. Much the same criticism was made in his discussion of my paper at the British Association meeting, 1910. Members of the British Psychological Society, whose recollection goes back to the pre-war years of 1909-13, will also recall how often, at its meetings, the validity of the various versions of this correction-formula supplied a topic for debate between Spearman and Brown (at that time the most active champion of the Pearsonian view: cf. also (18), pp. 83-92).

"assumes a knowledge of just the very figures" (the saturation coefficients) "you are attempting to calculate." "A circular procedure of this kind," he considered, was particularly unfair in any attempt to demonstrate the presence or absence of hierarchical arrangement, since "it is bound to exaggerate deviations from a uniform gradation"; and, "for the sake of argument," he tentatively put forward a substitute of his own, "to show that the assumption could be evaded."¹ In his own researches, however, he himself seems never to have used either this or any similar formula for calculating saturations: for all such purposes, including the determination of group factors, we needed, so he maintained, external or independent criteria, i.e., what he later called 'reference abilities' (cf. 26, p. 223).

To say therefore that "the calculation of the *g* saturation of each test formed an important part of Spearman's process" (32, p. 153) appears quite mistaken.² Occasionally research students, who were working with us jointly, calculated 'saturations' by my formula, or some equivalent device, as a supplementary procedure. But it will be noted that, in summarizing their work or my own, Spearman himself never quotes such figures.

The metaphor of 'saturation' is taken from Galton; and to account for the squaring Spearman refers to Galton's fraction which I have already quoted on p. 155 above. But, unlike Galton's expression, Spearman's correlation is here not a 'hypothetical correlation' (except for the correction): it is, as he insists, 'empirically observed.'

In his book there is no reference to saturation coefficients in the index, nor are any given in the text. Where, for example, he is discussing 'the amount of *g*' (chapter XII), or quoting the results of my own research-students or myself, we might certainly have expected the comparisons to be based on factor-saturations. Yet, throughout the chapter, the 'correlations with *g*' refer to *external* criteria, such as teachers' assessments, Miss Davey's "two standards of reference," or "the Binet-Stanford test series which served as a standard of reference." On p. 218 he prints two correlation tables for 'normal' and 'defective' children (similar to the tables I had given in making the same comparison in (23), Tables XVIII and XXIV); but no factor-saturations are computed. I have no wish to deprecate the line of approach that he has adopted. On the contrary, it appears to me a most valuable procedure which to-day has fallen undeservedly into neglect. My only point is that it does not constitute 'factor analysis' as now generally understood, i.e., a wholly internal analysis, as distinct from an external comparison with independent reference-values.

Spearman's Later Criteria. So far I have illustrated the difference between the two points of view by applying my procedure to Spearman's table. The difference will become plainer still if we glance for a moment at the methods Spearman proposed to substitute in examining my own tables.

1. In his article on General 'Ability' (19, p. 61) he reprints my 13×13 table in full; and then, instead of looking for isolated instances of enlarged coefficients, he calculates the

¹ My own contention had been that "in disproving Pearson's view" (about the presence of disproportionately high coefficients) "we ought at least to apply Pearson's own principles." To this Spearman replied: "accepting [this] for the sake of argument, the formula used should still be one that does not presuppose a knowledge of an unknown figure," and accordingly suggested "for trial . . . an alternative to your own equation, which does not suffer from this defect. But that does not mean I admit the principle." However, as is plain from the proofs, both this and his later substitute assume that we are dealing with tables that are strictly hierarchical: unlike my own formula, they yield inaccurate values for the saturations, when group factors are involved.

² It is true that on p. 276 of his article (12) Spearman brings together the theoretical "correlations with intelligence," obtained from his main groups and calculated by "eliminating observational errors"; and adds that their squares will indicate the "full absolute saturation of each activity with General Intelligence." The wording has apparently led later readers to suppose that Spearman was here printing 'saturation coefficients' in the sense in which the phrase is *now* employed in current writings, i.e., what I called 'estimated correlations with the hypothetical common factor.' But the context makes it clear that he is using for his criterion the external evidence derived from intelligence-ratings or examination-lists.

intercolumnar correlations for the "5 smallest" and "4 largest pairs of columns." On his hypothesis each correlation should be unity (with an allowance for sampling errors). The average values actually found come to no more than 0.59 and 0.81 respectively; but they are raised to 0.67 and 1.14 by correction. In his book he reverts to the same table, and gives the *mean* intercolumnar correlation as 1.06 (26, p. 139). As regards the figures themselves, it should be noted (i) that they are based on only 9 out of 78 possible correlations, (ii) that, by taking the mean correlation, an impossibly high value (1.14 and over) is made to compensate for the low, and (iii) that a favourable conclusion is only reached if we accept the *corrected* coefficients: if we take the raw correlations as calculated, the low values are more in keeping with my hypothesis than his. As regards the general principle involved, the difference between our assumptions is brought out most conclusively by Maxwell Garnett, himself formerly a lecturer in Karl Pearson's department. Starting, as Pearson did, by treating the observed test-measurements as weighted sums of the factor-measurements, he has shown, by a rigorous algebraic proof, that, "when Mr. Burt's conditions for a hierarchy are satisfied, the n qualities tested can be expressed in terms of $n + 1$ independent factors, viz., one general and n specific factors": whereas, when "Professor Spearman's correlation-between-columns conditions are satisfied," this result does not necessarily follow.¹

2. This criticism Spearman appears tacitly to have accepted; and, to meet it, he put forward the well-known 'tetrad-difference' criterion (24, 25).² Rigorously interpreted, the criterion would imply that only a proportional or 'synclinal' table could now be accepted as strictly 'hierarchical.' In his book, however, except in the chapter illustrating the alleged contrast between correlations for physical and mental measurements, the new criterion is not applied to the *individual* correlations.

To illustrate his procedure, may I again take his analysis of my own correlation table (26, pp. 234-5)? In my 1909 research (14, p. 164) I had claimed that there was some slight evidence indicating a special factor for sensory discrimination underlying the visual, tactile, and kinæsthetic tests—evidence that was later confirmed with larger groups. Spearman, on the other hand, argued that my figures really corroborated his view,³ namely, that the observed correlations could be "wholly explained by g ." My own procedure had been to test each individual deviation by its probable error. The principle of summation, however, permits a less laborious, if somewhat less informative, approach. Using a group factor analysis, we can partition the entire correlation matrix into m^2 submatrices— m being the number of suspected group factors. In the table in question there are at least three—motor, sensory, and memory. It will be seen that this device assumes that the empirical matrix of observed correlations may be regarded as formed by the superposition of a series of theoretical matrices of rank one—namely, a set of smaller hierarchies, each limited to a single group of tests, superimposed upon one large hierarchy, resulting from the factor common to all the tests. Provided that the group factors do not overlap, we can then 'condense' the table of correlations into an $m \times m$ matrix; and in this way each cluster of augmented correlations can be compressed

¹ *Brit. J. Psych.*, IX, 1919, p. 356 and refs.

² The first to suggest such a criterion—the usual determinantal criterion for a matrix of 'rank one'—was W. F. Sheppard at a British Association discussion on factors (1920). The term 'tetrad' was taken from Yule's discussion of the contingency table in cases of manifold association (*loc. cit.*, pp. 68-69). Spearman's own important contribution was the elaboration of an ingenious formula for the probable error (25, 26, p. xi).

³ In this paper I am concerned only with methods, not with conclusions. However, to understand Spearman's change of approach it should be remembered that his ideas on the concrete nature of " g " had by this date undergone a radical change. He had now accepted my demonstration that 'general intelligence' could *not* be identified with general sensory discrimination, but seemed rather to be manifested most fully in higher cognitive processes, such as those requiring "the perception, implicit or explicit, of a relation, and the reconstruction of an analogous one by 'relative suggestion'" (17, p. 101), or, as Spearman later called it, the 'eduction of relations and correlates' (26, pp. 165, 205). At the same time g was also identified with a general cognitive energy; and thus in some measure the views and criticisms of Wundt were also accepted: "es mag sein," Wundt had written, "das die *Energie* der Aufmerksamkeit den Zentralfaktor abgibt, der die einzelnen Tätigkeiten bestimmt" (7, 6th ed., p. 598): see 19, p. 69, where Spearman explains my high saturation coefficients for tests of 'attention' in much the same way.

into a diagonal cell.¹ Spearman's method is commonly described² as though it followed the same principle: but a closer study of his own account shows that this is by no means correct. His method obliges him to reduce the entire correlation matrix in every case to a single tetrad. Thus, in dealing with my own table, after separating off the three discrimination tests, he treats all the rest as pure tests of *g*, and uses them as 'reference values.' Putting unity in the diagonal (as Pearson's method likewise requires), and adopting a modified version of the formula for calculating the correlation between sums, he then proceeds to 'estimate the correlation between the two pools.' A single tetrad-difference is finally evaluated from these *estimated* correlations, and found to be non-significant. His figures for my own table are reproduced below in Table VIIA.

TABLE VII. THE TETRAD DIFFERENCE CRITERION

Groups	A. SPEARMAN (Correlations between Pools)		B. BURT (Average correlations)	
	Other Tests	Sensory Discrimination	Other Tests	Sensory Discrimination
Other Tests	·813	·386	·467	·154
Sensory Discrimination	·386	·167	·154	·251
Determinant ('Tetrad Difference')	-·013		+·093	

Ingenious as it is, this device does not really answer the question he himself has raised, namely: is one factor alone sufficient to account for the correlations, or, in other words, does the entire matrix form a matrix of rank one? Where this is the issue to be decided, we should insert in the diagonal, not unity, but the estimated self-correlations. The 'condensation' must then be effected by simply summing or averaging the coefficients as they stand. If the original 13×13 table is of rank one, the 2×2 table of sums or averages will also be of rank one, and its determinant will vanish. The figures so obtained are shown in Table VIIb. It will be seen that Spearman's procedure has greatly diminished the true size of the determinant.

Note too that, even when a significant departure from a simple 'hierarchical' arrangement has been proved in this way, Spearman's procedure does not allow us to decide where the cause lies: the positive value of the tetrad difference may be due to a group factor present either (i) in the discrimination tests, or (ii) in the other set of tests, or (iii) in both the sets. The fact that Spearman *calls* the remaining set of tests 'reference values for *g*' does not suffice to *prove* that they contain nothing but *g*: on the contrary, they almost certainly contain a 'learning' factor and probably a 'motor' factor as well.³

¹ It has been stated that "Spearman's system . . . looks upon any complex matrix of correlations as being due to lesser hierarchies superimposed upon the hierarchy" (32, p. 242). This hardly seems a correct account of Spearman's procedure. Indeed, his aim is rather to show that the assumption of 'lesser hierarchies' is not right, but wrong; and, on the rare occasions when he does admit a supplementary factor, his calculations treat both factors as oblique. (The statement quoted might be fairly applied to Pearson's equation for physical measurements, since Pearson's components are necessarily orthogonal; but the principle was first explicitly used in extracting a bipolar factor for emotional assessments: (21). For a worked example, cf. 'Factor Analysis by Submatrices,' *J. Psych.*, VI, Tables on pp. 349 and 352.)

² Admittedly the brief explanation given in his book (26, p. 223, footnote, and appendix, pp. xxi-xxii) is by no means clear. The account given above, together with the tables, are from résumés circulated beforehand for a debate opened by Spearman on 'The Nature of Intelligence' (British Psychological Society, Education Section, Dec., 1924). In his book (p. 235) 'auditory' is apparently a slip for 'tactile' (given in the résumé: the 'auditory' test was grouped with the tests for *g*). In his later applications of the method Spearman usually divided his 'reference abilities' (as well as the tests to be investigated) into two more or less equally balanced sets, so that the resulting 2×2 matrix (or 'tetrad') is no longer perfectly symmetrical: but this introduces no new principle.

³ One of many striking examples is to be seen in his examination of Rogers' work (26, p. 230). There verbal tests of Reading, Completion, Opposites, etc., are treated as pure tests of *g*, and the group factor is assigned to the arithmetical tests. But we could just as well assign a group factor to the verbal tests.

VII. THE THEORY OF TWO FACTORS AS A SIMPLIFICATION OF THE THEORY OF MULTIPLE FACTORS

The Three-Factor Theory. As I have already noted, in my own earlier researches the results strongly suggested "a small but discernible tendency for groups of allied tests to correlate together," even after the general factor had been removed; and the existence of such 'group factors' was decisively confirmed when Mr. Moore and I adopted the method of 'group testing,' and so were able to secure data from larger numbers.

This conclusion led to a working classification of abilities into three distinguishable types according to their range: (i) a general ability entering into *every* test belonging to a certain broad *genus*; (ii) special abilities, each limited to certain groups or *species*; and (iii) individual or specific abilities, each peculiar to a *single* test. The distinctions, which were relative rather than absolute, were originally put forward on *a priori* grounds¹; but appeared gradually to secure empirical verification. Later work in London schools demonstrated beyond all question that, in the cognitive and the emotional fields as in the physical, we had to reckon with "a multiplicity of common factors"; and that the appropriate factorial technique would need to be a modification of Pearson's method of "multiple correlation" (cf. 21, 22, and 23, pp. 53-6).

The classification proposed was subsequently styled, not perhaps quite accurately, a 'three-factor theory.' No new 'theory' was intended. The purpose was merely to define more precisely, in correlational terms, the distinctions that had already been implicitly assumed, both by writers on individual psychology like Galton and Binet and by most educationists and teachers.²

The Two-Factor Theory. In his first paper, as we have seen, Spearman began with the broader twofold contrast, repeatedly drawn by Galton, Wissler, and others, between 'general ability' on the one hand and 'special abilities' on the other. The novel feature in his work was not (as is so often supposed) the concept of 'general ability,' but his contention that what previous writers had called 'special abilities' were in point of fact "vanishingly minute" in their range. "The specific factor," he declared, "seems in every instance to be new and wholly different from that in all the others." Thus 'specific abilities,' in the sense of factors supposed to be characteristic of a broad *species* of cognitive tests, became reduced to 'specific factors' in the narrow sense of factors specific or peculiar to a *single* test.

As he rightly emphasized, such a conclusion, if true, would introduce a profound simplification into the whole science of human ability. To contrast it more clearly with the pluralistic theories of earlier writers, Sancti de Sanctis christened it a "theory of two factors."³ Taking the phrase for his title, Spearman in 1914 published the first explicit formulation of his famous two-factor hypothesis.⁴

¹ I regarded it as an *a priori* hypothesis, because, antecedently to any experimental work, it could be assumed that the possible conclusions found could always be expressed in terms of three types of proposition: (i) 'all cognitive processes contain this factor'; (ii) 'some cognitive processes contain this factor'; (iii) 'this cognitive process alone contains this factor.' Accordingly, borrowing the terms of traditional logic, I suggested calling them 'universal,' 'particular,' and 'singular' abilities respectively. (The Oxonian nomenclature was due to my former tutor in Logic, Dr. H. Hughes, to whom I was much indebted for help in clarifying the logical issues involved.)

² The threefold classification was first suggested in a paper given to the Manchester Child Study Society in October, 1909 (cf. 16, p. 95). It was later extended to affective and conative factors (cf. 'General and Specific Factors underlying the Primary Emotions,' *Brit. Ass. Ann. Rep.*, LXXXV, 1915, pp. 694f.).

³ *Proc. Internat. Cong. Medicine* (London, 1913), Section for Psychiatry.

⁴ 20, 24. A less sweeping form of the same theory had been put forward by Hart and Spearman in their joint paper (19).

It has of late been frequently asserted that the more comprehensive notion of a multiplicity of factors arose as a subsequent amendment or extension of Spearman's theory of two factors (e.g., Thurstone, 35, p. vi, 273). Exactly the opposite was the case. Spearman himself constantly pointed out that the older views of mental abilities, like those of Galton, Wissler, and Thorndike, implied a "multifocal theory"; and, when introducing his own simpler hypothesis, argued that mental testing would gain, "both in purpose and in method," if these multifocal theories were finally "abandoned" in favour of the hypothesis of a single universal factor, supplemented of course by the unique or "specific factors" which every theory was compelled to assume.

The way in which the older multi-factor theory may be reduced to a two-factor theory is shown most clearly in one of his later articles, where he asks: "What then are the fundamental qualities involved in measurements of abilities and their correlations?" In accordance with the 'general practice,' he says, we may begin by "writing the value of each ability in the following well-known form:"

$$x = w_1e_1 + w_2e_2 + \dots + w_ne_n,$$

where the e 's denote the independent 'elements' (or factor-measurements) and the w 's the coefficients (or weights). He then discusses the most plausible number of elements (n). They may be (i) "extremely numerous," (ii) "moderately numerous," or (iii) "comparatively few." But each of these alternatives still involves awkward consequences, when we come to deal with correlations between *psychological* characteristics. There remains, he says, (iv) a fourth possibility. "To obviate the difficulties we have only to accept the doctrine of two factors, and suppose that for each ability only *two* of the coefficients take values other than zero, one supplying an element specific to that ability alone, whilst the other gives an element common to all abilities."¹ The solution, it will be observed, is tantamount to the exclusion of all supplementary factors, whether group or bipolar.

Spearman's initial equation is the same in form as Pearson's equation for expressing correlated measurements for physical traits in terms of uncorrelated components. Let us consider, therefore, how Spearman's proposed simplification will affect the expected correlations. As we have seen, Pearson had asserted that, owing to the unity of the individual organism, even *dissimilar* organs and functions would show a moderate degree of positive correlation, and that *similar* organs or functions would show an augmented degree of correlation. Spearman draws a sharp line between physical and mental characteristics. Where mental characteristics are concerned, he accepts the first proposition, but denies the second. I myself, for example, had maintained that, "formally at least, the factor-structure obtained from a set of correlations for bodily measurements shows little difference from that obtained for cognitive or emotional assessments." Spearman, on the contrary, declared that, judged by his own criterion, "more discrepant distributions . . . could hardly be conceived." "In the mental field one factor alone accounts for all observable correlation." "And, having surveyed the whole mass of accessible data," he concluded that the published tables "all agree with our aforesaid principle of hierarchical arrangement": "the strikingness of the agreement, by contrast with the non-mental correlations, which many psychologists had likened to mental ones, was wonderful."² In short, whereas Pearson's formula tacitly assumed that his correlation matrices would in general have a rank of n (n being the number of traits), Spearman in effect contended that, since cognitive activities were governed by one common cognitive factor only, all the tables obtained with cognitive tests would have a rank of one and no more.

The confirmation of this hypothesis, he claimed, was "equivalent to a Copernican revolution." "Instead of blindly accepting a multiplicity of aimless wandering factors, all human ability is seen to revolve round a single central source of energy, the universal g ."

The Existence of Group Factors. The subsequent developments³ of Spearman's theory have been fully discussed elsewhere; and little need be added here. Enough, I hope, has now been said to dispel the notion, which seems of late to have acquired a widespread currency, that 'British factorists' had almost unanimously accepted a two-factor theory (cf. 35, pp. vi, 273). Many of us were prepared to give him our

¹ 'The Sub-structure of the Mind,' *Brit. J. Psych.*, XVIII, 1928, pp. 250f. I have substituted w_1 , etc., for Spearman's x_1 , etc., since to use the same letter, alike for the measurements and for the weights, is needlessly confusing. The various 'cases' considered roughly represent the views of (i) Thomson, (ii) Pearson, (iii) myself, (iv) Spearman (after 1920).

² 'Pearson's Contribution to the Theory of Two Factors,' *Brit. J. Psych.*, XIX, p. 96: cf. (26), p. 142.

³ A complete bibliography of Spearman's writings will be found in 34, pp. 382-5.

warm support so long as he was merely content to claim, with Galton and Binet, that general ability had far more influence than the more specialized abilities ; and readily took his side against the more uncompromising critics who declared that there is no such thing as general ability.¹ On the other hand, when Spearman himself proceeded to the opposite extreme, and declared that there are no such things as group factors, he ceased to carry with him any but the most ardent of the enthusiastic 'unitarians.'

Spearman's later work, like much of his book on *The Abilities of Man*, is concerned with a critical examination of the evidence for the alleged group factors. "The main upshot," he concludes, "is negative" (p. 241). Here the divergence between us becomes still more marked. Summarizing the researches available at that time, I had stated² that at least a dozen group factors had already been established, and that still more seemed on the verge of confirmation : reviewing much the same evidence, but taking a severer criterion of significance (5 p.e. instead of 3 p.e.), Spearman himself could find no more than "four exceptional cases" ; and these, he held, were due rather "to past experience than to native aptitude." Later he slightly increased the number. But on the whole, as Thurstone observes, he and his followers showed a marked tendency to reject erratic traits and tests as mere "disturbers of the tetrad equation," instead of re-examining them and exploring any new possibilities which their presence seemed to open up (35, p. 273).

Evidently, whether we belittle their number or reject them as mere "disturbers" of the main hypothesis, such irrepressible cases are fatal to "the grand simplicity of the two-factor theory." By way of defence, Spearman, in his later writings, insisted that the complete formulation of the two-factor theory had always included a stipulation that the battery should not contain tests or traits that are closely allied or similar : obviously, he adds, by taking two or more traits that are sufficiently alike, anyone can succeed in augmenting the resulting correlations almost as much as he wishes (cf. 26, pp. 103-4). But any such proviso plainly involves an abandonment of the two-factor theory as such. It asserts no more than the obvious truism that there are only two factors except when there are more ; and, to say the least, it marks a wide departure from the original announcement that "the specific element seems in every case to be wholly different from that in all the others."³

Once this reservation is made an essential part of the theory, Spearman's case against the older and more eclectic view breaks down : for that view had also explained the presence of augmented correlations by attributing them to traits that were (in Pearson's phraseology) 'like' rather than 'unlike.' Indeed, as Kelley, McDougall and other critics pointed out, the qualifying clause in the fuller formulation really covers a retreat to the old multi-factor position, and the new claim becomes "impregnable."⁴

Spearman's picture of the mind as a vast collection of highly specialized mechanisms, all operated by a single form of 'psychoneural energy,' formed an attractive speculation ; and the controversies which it provoked supplied a most valuable stimulus to sustained research. During the earlier stages Spearman was undoubtedly correct in insisting that the introduction of any additional factor should not be countenanced except when the statistical evidence was conclusively in its favour. Yet, to doubt the existence of this or that particular group factor, it is by no means necessary to doubt the existence of all group factors as such. Even at the outset it appeared a foregone conclusion that, by choosing our traits appropriately and by increasing the size of our sample until the probable errors are sufficiently small, we could always discover some sort of supplementary factors in addition to the first.

¹ Cf. Board of Education, *Report on Tests of Educable Capacity*, 1924, p. 236. P. 19 expresses the views of those who held a 'three-factor theory' ; appendix IX summarizes the conflicting opinions prevailing at that date.

² *The Measurement of Mental Capacities*, pp. 31f. and refs.

³ See 12, 'Summary of Conclusions,' p. 284 ; my italics ; Spearman italicized the whole of this conclusion.

⁴ Cf. 27, p. 37, and 30, p. 96.

Hence (if I may repeat what I have said on an earlier occasion) "in my view, the real question to be determined is not: Are there or are there not any supplementary factors? but: What is the relative importance of these supplementary factors and the general factor respectively? Spearman's method does not enable us to answer this broader question. But if we adopt a summational procedure, and apply it to each correlation table as a whole, we can, in principle at least, always hope for a clear and decisive answer; and, whenever Spearman's special problem becomes relevant, we can solve that incidentally by the same general approach."

I think, therefore, we may fairly conclude that, notwithstanding Spearman's criticisms, "the Galton-Pearson school" has after all "provided the model" or at least the main line of development, for factorial work in psychology. From a mathematical standpoint, the methods of factor analysis in vogue at the present time resemble in their general approach, not so much the somewhat specialized technique which Spearman proposed on the basis of his own somewhat specialized hypotheses, but rather the older procedures, first outlined by Edgeworth and Karl Pearson, for reducing correlated variables to uncorrelated components. Founded on these more general principles, they have proved applicable, not only to psychological work, but to numerous other fields besides. To-day, I imagine, Spearman's 'general' and 'specific' factors would be almost universally regarded as merely extreme and limiting cases of 'factors' less narrowly defined. As such it was natural for them to attract attention first, and receive disproportionate emphasis; but they can no longer be regarded as the only factors entering into the structure of the mind.

VIII. SUMMARY AND CONCLUSIONS

1. Spearman's 'two-factor theory' was derived quite as much from current psychological assumptions as from statistical data, and must therefore be examined from the former standpoint quite as much as from the latter. When he commenced his investigations into mental abilities, the theory of a single cognitive function, advocated by such different writers as Spencer, Ward, and their respective followers, was beginning to supersede the theory of multiple faculties, still held by individual psychologists like Binet and Galton.

2. Criticizing the earlier correlational work of Wissler, Thorndike, and other members of the 'Galton-Pearson school,' Spearman contended that, with an improved statistical procedure, it would be possible to prove the all-sufficiency of what Galton had termed 'general ability' and the absence of anything like 'special abilities' in the original sense of that phrase. Of the two main 'sub-theories' as to the nature of the single fundamental function, Spearman inclined to that of Spencer and Bain (which identified it with 'discrimination') rather than to that of Ward and Wundt (which identified it with 'attention' or 'apperception').

3. The novel feature in his procedure consisted, not in 'factor analysis' as now understood, but rather in a method which has a closer affinity with so-called 'canonical analysis.' Instead of seeking factors by an internal analysis of the correlation table as a whole (as Pearson and others had proposed), he endeavoured, by a process of 'correction,' to determine the maximum correlation between a combination of two or more tests on the one hand and a combination of two or more external reference-values on the other. Supplementary criteria were devised to corroborate the absence of group factors.

4. The novel feature in his conclusions was his attempt to demonstrate that so-called 'special abilities' were wholly specific. As a result he argued that the substitution of a 'unifocal' hypothesis for the older 'multifocal' hypothesis would effect a valuable simplification in the theory of cognitive abilities. It is the attractive simplicity of this hypothesis, rather than the factual or statistical evidence, that accounts for its continued popularity with many British writers.

5. However, Spearman's own correlations, when factorized as a whole *without* applying 'corrections' or 'eliminations,' are found to disprove any contention that such tables can be regarded as matrices of rank one, and so explained in terms of a single common factor only. And the subsequent development of psychological theory, both on the statistical and the non-statistical sides, has tended to confirm the earlier views and methods of the Galton-Pearson school rather than those of Spearman. His own eventual admission of at least occasional group-factors reintroduces the multiple-factor standpoint, and makes the tetrad-difference criterion obsolete. The results obtained by multiple analysis support the view of a hierarchical structure of the mind, dependent not on a single general factor only, but on group factors of various levels and varying degrees of generality.

REFERENCES

1. Galton, F. (1869). *Hereditary Genius*. London : Macmillan.
2. Galton, F. (1889). *Natural Inheritance*. London : Macmillan.
3. Sully, J. (1892). *The Human Mind : A Textbook of Psychology*. London : Longmans.
4. Binet, A. (1900). 'Attention et adaptation.' *Ann. Psych.*, VI, 248-404.
5. Pearson, K. (1900). *Grammar of Science* (2nd ed.). London : Black.
6. Wissler, C. (1901). 'The correlation of mental and physical tests.' *Psych. Rev. Mon.*, III, 1-62.
7. Wundt, W. (1902). *Grundzüge der Physiologischen Psychologie* (5th ed.). Leipzig : Engelmann.
8. Thorndike, E. L., *et al.* (1902). 'Correlations among perceptive and associative processes.' *Psych. Rev.*, IX, 374-382.
9. Thorndike, E. L. (1903). 'Heredity and correlation in school abilities.' *Columbia Univ. Contributions*, XI, ii.
10. Binet, A. (1903). *L'étude expérimentale de l'intelligence*. Paris : Alcan.
11. Pearson, K. (1904). 'On the inheritance of mental and moral characters.' *Biometrika*, III, 131-190.
12. Spearman, C. (1904). (i) 'The proof and measurement of association between two things.' (ii) 'General intelligence objectively determined and measured.' *Amer. J. Psych.*, XV, 72-101, 202-293.
13. Krueger, F., and Spearman, C. (1906). 'Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten.' *Zeit. f. Psych.*, XLIV, 50-114.
14. Burt, C. (1909). 'Experimental tests of general intelligence.' *Brit. J. Psych.*, III, 94-177.
15. Spearman, C., Burt, C., Brown, W., Meumann, E., Myers, C. S. (1910). 'Mental tests.' *Brit. Ass. Ann. Rep.*, LXXIX, 804f.
16. Burt, C. (1911). 'The experimental study of general intelligence.' *Child Study*, IV, 33-44, 92-101.
17. Burt, C. (1911). 'Experimental tests of higher mental processes and their relation to general intelligence.' *J. Exp. Ped.*, I, 93-112.
18. Brown, W. (1911). *The Essentials of Mental Measurement*. (2nd ed., with Thomson, G. H., 1920.) Cambridge : University Press.
19. Hart, B., and Spearman, C. (1912). 'General ability : its existence and nature.' *Brit. J. Psych.*, V, 51-84.
20. Spearman, C. (1914). 'The theory of two factors.' *Psych. Rev.*, XI, 101-115.
21. Burt, C. (1915). 'General and specific factors underlying the primary emotions,' *Brit. Ass. Ann. Rep.*, LXXXIV, 694f.
22. Burt, C., and Bickersteth, M. (1916). 'Some results of mental and scholastic tests.' *Report of Conference of Educational Associations*, 30-37.
23. Burt, C. (1917). *The Distribution and Relations of Educational Abilities*. London : King.
24. Spearman, C. (1923). 'A further note on the theory of two factors.' *Brit. J. Psych.*, XIII, 266-270.

25. Spearman, C., and Holzinger, K. (1924). 'The sampling error in the theory of two factors.' *Brit. J. Psych.*, XV, 17-20.
26. Spearman, C. (1927). *The Abilities of Man*. London : Macmillan.
27. Kelley, T. L. (1928). *Crossroads in the Mind of Man*. Stanford : University Press.
28. Spearman, C. (1928). 'Pearson's contribution to the theory of two factors.' *Brit. J. Psych.*, XIX, 95-101.
29. Spearman, C. (1930). 'G and after : a school to end schools.' *Psychologies of 1930*. Massachusetts : Clark University Press.
30. McDougall, W. (1932). *Energies of Men* (Appendix to ch. VI, 'On the Application of the Method of Correlation'). London : Methuen.
31. Guilford, J. P. (1936). *Psychometric Methods*. New York : McGraw-Hill.
32. Thomson, G. H. (1939). *The Factorial Analysis of Human Ability*. London : University of London Press.
33. Spearman, C. (1946). 'The theory of a general factor.' *Brit. J. Psych.*, XXXVI, 117-130.
34. Thomson, G. H. (1947). 'Charles Spearman.' *Proc. Roy. Soc. (Obit. Not.)*, V, 373-385.
35. Thurstone, L. L. (1947). *Multiple Factor Analysis*. Chicago : University Press.
36. Meili, R. (1949). 'Sur la nature des facteurs de l'intelligence.' *Arch. Psych.*, XXXIV, 33-47.
37. Burt, C. (1949). 'Alternative methods of factor analysis and their relations to Pearson's method of "principal axes".' *Brit. J. Psych. (Stat. Sect.)*, II, 98-121.

THE RECIPROCITY PRINCIPLE AS AN AID TO FACTOR ANALYSIS

By JOSEPH SANDLER

Psychological Department, The Institute of Psychiatry, London

I. Introduction. II. An Extension of the Reciprocity Principle to the Singly-centred Matrix. III. Person-equivalents and Test-equivalents : (a) The Person-equivalent of a Test ; (b) The Test-equivalent of a Person. IV. Equivalents as aids to Rotation : (a) The Analysis of Persons ; (b) The Analysis of Tests. V. A Worked Example.

I. INTRODUCTION

It sometimes happens in psychological research that only a small number of persons is available for analysis. This is often the case in the clinical and orectic fields where problems of typology and classification arise. The clinical worker commonly seeks to overcome the difficulty by collecting as much data as possible about each case. A classical example is the elaborate typology set up by Freud (6) as a result of the detailed study of a comparatively small number of cases—a typology which subsequent experience appears to have confirmed (1, 9, 12).

For similar reasons the statistical approach need not be discarded any more than the clinical when the number of cases is small. Burt (3) and Stephenson (14) have shown that the factor analysis of persons provides a helpful technique when the sample of tests¹ is larger than the sample of persons. The equivalence of the factors obtained by factorizing correlations between persons to those obtained by factorizing correlations between tests has been demonstrated by Burt (4). His proof of what he calls the 'Reciprocity Principle' is perhaps the best argument for the acceptance of person factor analysis as a valuable factorial technique.

The Reciprocity Principle implies (with certain restrictions on the original scores) an equivalence or reciprocity between person factors and test factors.² It states that, provided appropriate units are used, the factor-measurements in the test equations are identical with the factor-saturations in the person equations, and the factor-saturations in the test equations with the factor-measurements in the personal equations.

To demonstrate *exact* equivalence he begins with a doubly-centred matrix of scores. He then proceeds to calculate correlations between the tests and covariances between the persons. The particular type of score matrix used by him thus forms a special case, and is extensively discussed by Thomson (15), who says (p. 218) : "Most of what we call the association or resemblance between either tests or persons, the amount of which we gauge by the correlation coefficient, is due to something over and above this [residual relation]. We can write down an infinity of possible raw matrices from which Burt's doubly-centred

¹ By *test* is meant any measurable attribute of a person.

² Except (as Burt explicitly points out) for the *general* factor found in either set of scores. The reduction to a 'doubly-centred' score-matrix is in fact due to the elimination of this first factor. Thomson's criticism is thus merely a restatement of Burt's own reservation. The first or general factor is necessarily independent of the others ; hence, if we seek to put back such a factor, we naturally find an "infinity" of ways in which it can be done. But the preliminary elimination of the factor, by taking deviations about the averages, is no more disadvantageous here than elsewhere.

matrix may have come. To the rows of the latter matrix we can add any quantity we like without in the slightest altering the correlations between the tests, but making enormous changes in the correlations between the persons . . . [and] we can add any quantities we like to the columns of the matrix of marks, and produce an infinity of different matrices of correlations between the tests. . . . It [Burt's theorem] does not apply to the full covariances of the matrix centred only one way, in the manner usually meant when we speak of covariances or correlations."

The first part of this paper attempts to meet this criticism, and show how an analysis of persons can be carried out with an initial score matrix that is centred *one* way only. This, of course, has already been done by numerous workers who have used Burt's technique. But here I propose to depart from the usual procedure of calculating either correlations or covariances between persons; instead the *product-sums* between persons will be calculated and factorized. In the second section I shall show that the saturations so obtained are directly related to the matrix of factor-measurements: this, if I understand him rightly is virtually the same result as has been found by Burt for the analysis of person covariances calculated from a doubly-centred score matrix.¹ In the following sections the idea of person-equivalents of tests and test-equivalents of persons will be put forward as a help in problems of rotating and identifying axes; and finally a worked example will be appended using rectangles for 'persons' and measurements of these rectangles as 'tests.'

II. AN EXTENSION OF THE RECIPROCITY PRINCIPLE TO THE SINGLY-CENTRED MATRIX

Consider the scores of N persons on n tests. Expressed in normalized form, so that the sum of the squares of the scores in any one test is equal to one, and the sum of these scores is equal to zero, they may be written as a score matrix S , the score of person i in test j being s_{ji} . Two specification equations can be written for s_{ji} , the first in terms of the test factors (hypothetical tests), and the second in terms of person factors (hypothetical persons). Adopting the notation used by Burt, we have from the analysis of *tests*:

$$S = F_t P_t, \quad (i)$$

where P_t denotes the factor-measurements for the persons, and F_t the test-saturations. From the analysis of *persons*, we have:

$$S' = F_p P_p \text{ or } S = P_p' F_p', \quad (ii)$$

where P_p' is an $n \times m$ matrix of scores of the m hypothetical persons (person factors) in the n tests, and F_p' is the $m \times N$ matrix of person factor-saturations for persons expressed in suitable units. Burt's Reciprocity Principle then states that, provided the factor-saturations have been normalized and thus expressed in suitable terms, F_t will be equivalent to P_p' and F_p' to P_t .

1. *The Analysis of Tests.* The correlation matrix R_t , containing the correlations between the tests, is given by: $R_t = SS'$ (iii), and R_t may be factored into a set of factors F_t , so that $R_t = F_t F_t'$. Assuming that F_t is the principal component solution, we have $F_t' F_t = V_t$, where V_t is a diagonal matrix.

2. *The Analysis of Persons.* Equation (iii) is a method of condensing S into a symmetrical matrix R of smaller order and easier to handle. This is advisable when $n < N$. When $n > N$, it is easier to take $R_p = S'S$, where R_p is the matrix of *person product-sums*. Substituting from (i) we have $R_p = (F_t P_t)' (F_t P_t) = P_t' V_t P_t = (V_t^{\frac{1}{2}} P_t)' (V_t^{\frac{1}{2}} P_t)$. Let $V_t^{\frac{1}{2}} P_t = F_p'$ (iv) then it is clear that R_p may be factored into a matrix F_p in the same way as R_t was

¹ The argument in Section II is equivalent to Burt's (4), except that the formal part of his proof was restricted to the calculation of person covariances from a doubly-centred matrix of scores. [On p. 40 of 4, however, it was expressed in terms of 'product-sums' or 'unaveraged covariances.' These become 'averaged covariances' if the tests are in 'unitary standard measure.' *Editorial Note.*]

factored into F . The matrix product $F_p F_p'$ reproduces the matrix of person product-sums R_p . F_p will be principal component solution, because it is the only solution which makes product $F_p' F_p$ yield a diagonal matrix. We have in fact $F_p' F_p = V_i^{-\frac{1}{2}} P_i P_i' V_i^{\frac{1}{2}} = V_i$. Since $P_i = V_i^{-\frac{1}{2}} F_p'$ (iva), the matrix of factor-measurements P_i may be derived directly from the factor analysis of the R_p matrix. When we analyse R_i , the factors are hypothetical tests, and the matrix F_i contains the saturations of the n tests with the m hypothetical tests or factors. When we analyse R_p we obtain a set of hypothetical persons, and the saturations of the N persons with the m person factors are written in F_p . The direct relation between the person factor-saturations and the matrix of factor-measurements P_i is thus given by equation iv or iva.

Any hypothetical person (principal-component factor) w , say, is saturated with one factor only, the factor w . His factor-saturation is equal to the square root of the factor-variance, and may be written as $v_{w\frac{1}{2}}$. Since the factors are independent, the saturation of person w with any of the remaining $m - 1$ factors is zero. Thus if the m factors are considered as m persons, the matrix of factor-saturations of these m persons with the m factors is a diagonal matrix $V^{\frac{1}{2}}$. The matrix of factor-measurements of the m persons may then be found from equation iva, as follows: $V^{-\frac{1}{2}} V^{\frac{1}{2}} = I$ (v). Once the factor-measurements of the m hypothetical persons have been found, the scores of these persons in the n hypothetical tests are given by equation ii: $P_p' = F_i I = F_i$.

The *Reciprocity Principle* requires that the matrix F_p' , which contains the factor-saturations of the persons, should be reduced to suitable units. If the matrix F_p' is normalized by premultiplying it by $V^{-\frac{1}{2}}$, we have $V^{-\frac{1}{2}} F_p' = P_i$, as required. From the last two equations it will be seen that the principle still holds when the singly-centred matrix of scores S is factored by persons and tests, provided that correlations for the tests and product-sums for the persons are the matrices analyzed.

III. PERSON-EQUIVALENTS AND TEST-EQUIVALENTS

My contention is that this extension of the Reciprocity Principle enables us to postulate certain relations between person factors and test factors, and under certain conditions to extend them to persons and tests in general. The person factors found from the analysis of R_p are the same as those obtained from the analysis of R_i . Yet in the first case the factors are hypothetical persons and in the second hypothetical tests. From equation (v) it is evident that the person factor w may be regarded as having a factor-measurement of 1 for factor w and 0 for each of the remaining $m - 1$ factors. Person w may conveniently be called the person equivalent of test factor w , and conversely test factor w is the test-equivalent of person w .

Any test t may be written as a linear combination of the test factors. This same combination of factors also yields a person, as the factors are at the same time hypothetical persons. The person-equivalent p_t of any test t has the property that his score on any other test k is a linear function of the correlation between t and k . This may be more easily seen from the geometrical picture.

The person factor matrix F_p' contains the saturations of the N persons with the m factors. These factors may be represented by axes at right angles to one another, and the persons as points or vectors in the factor space. If we normalize F_p' by equation iva, we have the factor-measurement space P_i . This is in effect a weighted or 'stretched' person factor space; and the factor axes are simultaneously persons and tests. If we can locate a test in this space, then this test-vector is also a person-vector. The direction of the test-equivalent p_t in the F_p' space is given by equation iv.¹ The direction of a person vector p

¹ Strictly speaking, the vector is collinear with a whole set of persons, all having the same direction, but varying in the sums of the squares of their scores on the n tests. For practical purposes the case can be limited to the one person in the P space whose vector is the same length as that of the test. If all the variance of the test t is accounted for by the factors, then this length is 1.

in the P_t space may be found from the saturations of p given in F_p' . Hence we can find the direction of the test-equivalent t_p of test p in space P .

(a) *The Person-equivalent of a Test.* To avoid an accumulation of subscripts let us now write F for F_t , G for F_p , R for R_t and Q for R_p . Given the person factor matrix G , and the scores $s_{t1}, s_{t2}, \dots, s_{tN}$ of the N persons¹ in test t , it is possible to determine the saturations of the person-equivalent p_t of test t . The saturation of t with any test factor w is given by

$$f_{tw} = r_{tw} = \sum_{i=1}^N s_{ti} p_{wi} = v_w^{-\frac{1}{2}} \sum_{i=1}^N s_{ti} g_{iw}. \quad (\text{vi})$$

The factor-saturations of test t fix the direction of the test vector t in the P_t space, and thus co-ordinates of the person-equivalent p_t are known. From equation (iv) we have

$$g_{tw} = v_w^{\frac{1}{2}} f_{tw} = v_w^{\frac{1}{2}} v_w^{-\frac{1}{2}} \sum_{i=1}^N s_{ti} g_{iw} = \sum_{i=1}^N s_{ti} g_{iw}. \quad (\text{vii})$$

Equation (vii) gives a simple working formula for finding g_{tw} , the saturation of p_t with factor w .

Thus the person-equivalent p_t has the property that his projection on any other test k is given by his score on test t multiplied by the correlation r_{tk} between tests t and k . Psychologically this is meaningful; and the person may be considered as the 'type' corresponding to a particular test. For example, we may describe an 'introvert' as a person who is the 'type' corresponding to a test of introversion t , even though we know that this quality is a combination of several factors (7). Conversely, the test-equivalent t_p is a measure of the resemblance of any person to person p .

The significance of the association between person p_t and the factors may be gauged by a simple application of the test for the significance of a multiple correlation coefficient R (17). The m factors may be considered as m uncorrelated tests with which test t has been correlated. The correlation of t with any factor w is given by equation vi. Thus the multiple correlation

R between t and the factors is given by $R^2 = \sum_{w=1}^m r_{tw}^2$.

An F-ratio tests the significance of R^2 , where $v_1 = m$, and $v_2 = N - m - 1$. R^2 is the proportion of the variance of test t accounted for by the m factors. This test may not be completely accurate when applied to the type of data discussed here as it assumes that the populations from which the scores on test t and the m hypothetical tests (factors) are drawn, are normally distributed.

(b) *The Test-equivalent of a Person.* Given a person p , the task is to find the direction of the test-equivalent t_p in the test factor space F . The available data consist of F and the scores of person p on the n tests.² The person product-sum between person p and any hypothetical person (factor) w , is given by $q_{pw} = \sum_{j=1}^n s_{jp} y_{jw} = \sum_{j=1}^n s_{jp} f_{jw}$. (viii)

q_{pw} is thus easily obtainable from the given data. As the saturation of hypothetical person w with any of the factors except w is zero, we may write $q_{pw} = g_{pw} g_{pw} = v_w^{\frac{1}{2}} g_{pw}$. The value g_{pw} , the saturation of person p with the factor w , is given by $g_{pw} = v_w^{-\frac{1}{2}} q_{pw}$. v_w is given by equation viii. If g_{pw} is known, then the direction of vector p in the G space may be found, and by equation (iva) we have $p_{wp} = v_w^{-\frac{1}{2}} g_{pw}$, and $p_{wp} = v_w^{-1} q_{pw}$. (ix)

Now the vector p in the P space is also the test-vector t_p . But the values of p_{wp} are proportional to the correlation of t_p with factor w taken as a test (i.e., to the saturation of t_p with factor w). Once the values $p_{1p}, p_{2p}, \dots, p_{mp}$ are known, they may be normalized to get the direction cosines of t_p in the F space. The vector corresponding to t_p in the P space may be found directly from the given data by the application of taking p_{wp}

$$= v_w^{-1} \sum_{j=1}^n s_{jp} f_{jw}.$$

¹ Normalized scores.

² These scores are written in the same units as the 'normalized' scores in S .

IV. EQUIVALENTS AS AIDS TO ROTATION

My chief purpose here is to suggest that the use of person- and test-equivalents may serve as guides to rotation. Tests may be introduced into the person factor space in the shape of their person-equivalents, and similarly persons may be introduced into the test factor space by finding the test-equivalents of these persons. Although the use of a test-equivalent of a person is proposed, this one person can be the most typical or the most representative person of a group of persons, i.e., the average person of that group. In practice the test-equivalents of such typical or average persons will commonly be used in preference to single individuals.

(a) *The Analysis of Persons.*—The advantages and disadvantages of factoring persons have been fully discussed elsewhere (3, 4, 14). Where the number of persons is small, the factor analysis of person product-sums is a useful technique. In such a case, as Burt has shown in the papers cited, the factors obtained are the same as would be reached by factorizing correlations between tests.

Any method of factorizing the matrix R_p , which is dependent on the mathematical properties of R_p , yields factors which are arbitrary in so far as psychological meaning is concerned, in that the mathematically determined factors are functions of the particular set of persons and tests chosen for analysis. In analyzing tests the arguments for the rotation of the axes to make them psychologically meaningful, and constant from analysis to analysis, have been given in full by Thurstone (16), Reyburn and Taylor (11), and Thomson (15); and the same arguments apply with equal force to the analysis of persons.

The application of the simple structure concept (16) to person factors is more difficult than in the case of tests. As Reyburn and Taylor point out (and this is conceded by Thurstone), it is necessary to put simple structure into the battery in order to find it. In other words, careful choice of the battery is necessary, and a certain number of tests of pure factorial composition should be used. With persons however, the disadvantage exists that it is impossible to have a known 'pure' person for future researches. Persons change; and in any case it is impossible to obtain duplicates, whereas tests are relatively permanent; many copies can be made; and they are easily transportable. On the other hand, the person factor space may be enriched by the person-equivalents of tests; and, if sufficient equivalent persons are used, we may get to know a great deal about the psychological nature of the factors in the factor space F_p .

If the suggestion of Reyburn and Taylor is followed, and we proceed with rotation by the formulation and testing of an hypothesis as to the presence or absence of a particular factor, then rotation may be aided by the insertion of appropriate person-equivalents into the person factor space. Suppose that a particular domain has been investigated by the analysis of person product-sums; examination of the person points in the factor space leads us to suppose that a certain factor A is present. We may test the hypothesis by introducing the person-equivalent of a test which has been shown in other researches to be a good measure of factor A ; or a set of tests which are all measuring A , may be used.

This procedure is made possible by the fact that a small group of persons can, as a rule, be fairly easily reassembled for further testing. The person-equivalent P_i should fulfil the following requirements if it is to be used as a confirmation of the hypothesis.

(1) The value of R^2 must be significant. The test must be significantly associated with the empirically determined factors.

(2) R^2 should be large in relation to the reliability of test t .

(3) The direction of p_i in the factor space G and the projection of the N person points on the vector p_i must be in accordance with the hypothesis.

Requirement two may be modified if it is known that test t measures some factor or factors other than A ; but it is necessary that these other factors lie outside the domain being investigated. The presence of these factors will reduce R^2 but the remaining requirements should still be fulfilled.

Bartlett (2) points out that the chief value of psychological typology lies in the demonstration that independent classifications, such as those found through factor

analysis, are correlated. If, for example, persons have been classified according to physique and temperament, a significant correlation between two systems would indicate the presence of an underlying 'factor.' If a set of persons has been factorized for a population of temperamental traits, measures of their physique may be introduced into the person factor space; and, if sufficient significant correlations are found, the rotational problem becomes much easier.

As an example an investigation undertaken by the writer (13) may be cited. 50 mental patients were factorized for content-items in the Rorschach test. Besides a general factor of 'Productivity,' a bipolar factor of 'Anatomical vs. Human-and-Animal' responses was found. The same 50 persons were then factorized for symptoms; and seven factors extracted. The second Rorschach factor was treated as a test; and its person-equivalent introduced into the person factor space. A significant R of 0.56 was found; and the vector p_i was seen to discriminate between (a) the excited, irritable, or paranoid patients and (b) the retarded or depressive patients. Since the two classifications were correlated, we may infer that the same factor is common to both systems

(b) *The Analysis of Tests.*—In the search for *simple structure* the test-equivalents of persons may play a most useful role. The smaller the number of tests in the factor space, the easier it is to find a simple structure. On the other hand, if the number of points in the factor space is large, there is a greater probability that the simple structure will be unique. The calculation of R involves a prohibitive amount of labour when the number of tests exceeds 50 or 60. In any case the tests should be as factorially pure as possible; for, as Thurstone (16, p. 333) points out, "tests of high complexity . . . cannot be expected to contribute to the identification of primary factors."

A set of factorially pure persons of low complexity (well-known 'types') can be more readily selected, and the test-equivalents of these persons can then be added to the test factor space. Alternatively, once simple structure has been found from the test projections, the psychological identifications of factors may be confirmed by choosing an appropriate set of persons.

Thurstone (16, p. xii) observes that the study of 'freaks' may be "much more likely to be revealing of the underlying nature of the domain than are carefully randomized samples from the general population"; and this suggestion is well worth following in the analysis of persons.

Further use of test-equivalents may be illustrated by considering the domains of temperament and personality. Kretschmer (10), it will be remembered, has divided normals into two broad groups of cyclothymes and schizothymes; and Eysenck (5) has used the classification of neurotics into hysterics and dysthymics, found from the factor analysis of neurotic subjects, as a basis for generalizations about dimensions of the normal personality. Now, if non-neurotic persons are being studied, then psychotics and severe psychoneurotics correspond to 'freaks.' Here therefore is a possible method for testing the hypothesis that the normal personality can be described in terms of psychiatric classifications.

V. A WORKED EXAMPLE

In the following example, five rectangles will be used in place of persons. The two dimensions or 'factors' are assumed to be present, viz., height and breadth, and the figures for these are given in Table I. The correlation between the two is 0.40.

TABLE I. THE FACTOR MEASUREMENTS OF THE PERSONS

Persons	Dimensions	
	h	b
1	6	5
2	3	4
3	7	3
4	1	2
5	4	1

The 'tests' are constructed as follows: (1) $b + h$; (2) $3b - h$; (3) $4h - b$; (4) $7h + b$; (5) $5b + h$. The normalized scores are shown in Table II.

TABLE II. NORMALIZED TEST SCORES AND FACTOR SCORES

		Persons				
		1	2	3	4	5
Tests	1	·5677	—·0299	·4183	—·6275	—·3287
	2	·4793	·4793	—·3195	·0228	—·6618
	3	·2876	—·3208	·6196	—·6527	·0664
	4	·4195	—·2126	·5631	—·6723	—·0997
	5	·6469	·2083	·1535	—·4495	—·5592
Factors	$\begin{cases} h \\ b \end{cases}$	·3770	—·2513	·5864	—·6702	—·0419
		·6325	·3162	·0000	—·3162	—·6325

(a) *Factorizing by Tests.* The matrix of correlations between the tests ($S'S$) is shown in Table III. The columns headed I and II give the saturations with the two principal components

TABLE III. THE CORRELATIONS BETWEEN TESTS AND PRINCIPAL COMPONENTS

Tests	1	2	3	4	5	I	II
1	1·000	·327	·820	·934	·891	·9989	—·0463
2		1·000	—·273	—·031	·721	·2831	—·9594
3			1·000	·970	·471	·8455	·5339
4				1·000	·670	·9496	·3125
5					1·000	·8691	—·4952
Square-Sum						3·4499	1·5505

Once the tests have been factorized, it is possible to introduce the test-equivalents of any persons we choose into the test factor space. Suppose that an investigator finds two types of person predominant, the *tall*, and the *broad*. He then introduces a typical specimen of each type into the test factor space. If the representative 'tall' person is taken arbitrarily to be of height 10, his breadth, if he is really typical, will work out to be 6·60. Similarly the typical 'broad' person of breadth 7, will have a height of 4·53. The tall and broad persons will be denoted by H and B respectively. The scores of H and B in the 5 tests, written in the same units as the scores in matrix S , are as follows :

	Tests				
	1	2	3	4	5
H	1·1250	—·1381	1·1987	1·2104	·7373
B	·9562	1·0955	·3098	·5976	1·2280

The person product-sums of these two persons with factors I and II are as follows (equation viii) :

Person H : With Factor I, 3·8884 ; with Factor II, ·7335
 Person B : With Factor I, 3·1620 ; with Factor II, —1·3512

The co-ordinates of the test-equivalents of H and B in the test factor space, are given by equation ix, as-fol-lows :

t_H : With Factor I, 1·1271 ; with Factor II, ·9165
 t_B : With Factor I, ·4731 ; with Factor II, —·8715

If these two pairs of values are normalized, we have the direction cosines of the two test-equivalents t_H and t_B in space F .

t_H : With Factor I, ·9221 ; with Factor II, ·3870
 t_B : With Factor I, ·7247 ; with Factor II, —·6891

Now these values are the co-ordinates of two tests which are equivalent to two typical persons. The correlation between the two test-equivalents t_H and t_B may be found from these direction cosines, vi :

$$(\cdot 9221)(\cdot 7247) + (\cdot 3870)(-\cdot 6891) = \cdot 40$$

Thus the two 'typical' persons have been used to locate the factor axes in the test factor analysis.

(b) *Factorizing by Persons.* The product-sums for persons are calculated in the usual way. Table IV gives the results, as well as the saturations of the persons with the two principal component factors.

TABLE IV. THE MATRIX OF PERSON PRODUCT-SUMS AND PRINCIPAL COMPONENTS

Persons	1	2	3	4	5	I	II
1	1.229	.166	.598	-1.106	-.887	1.0261	-.4193
2		.422	-.452	.288	-.424	-.1005	-.6411
3			1.002	-1.122	-.026	.8184	.5765
4				1.474	.465	-1.1852	-.2636
5					.873	-.5589	.7486
Variance						3.4498	1.5491

The values in the bottom row (designated 'Variance') are the sums of the squares of the person factor-saturations. These are identical with h^2 in Table III, except for 'rounding-off.'

The two tests, h and b , may be introduced into the person factor space G in the form of their person-equivalents p_h and p_b , according to equation vii. This involves the multiplication of rows h and b in Table II by columns I and II in Table IV, yielding the saturations of p_h and p_b with factors I and II, viz.: (i) $g_{hI} = 1.7097$; $g_{bI} = 1.3455$; and (ii) $g_{hII} = .4864$; $g_{bII} = -.8581$. The position of the two points, p_h and p_b , in the person factor space may be compared with the position of persons H and B . The saturations of H and B with the person factors may be found by multiplying the values q_{hI} , q_{bI} , by v_I^{-1} , and q_{hII} , q_{bII} , by v_{II}^{-1} . The numerical values of q_{hI} , q_{bI} , q_{hII} , and q_{bII} have already been given. We thus find that the saturations of H and B with the person factors are as follows: (i) Person H : $g_{hI} = 2.0935$; $g_{hII} = .5893$; (ii) Person B : $g_{bI} = 1.7024$; $g_{bII} = -1.0856$. And if the points p_h , p_b , H and B are plotted in the person factor space, it will be immediately evident that each pair of points p_h and H , p_b and B , have the same direction in the factor space.

REFERENCES

1. Abraham, K. (1942). *Selected Papers*. London: Hogarth Press.
2. Bartlett, M. S. (1948). 'Internal and external factor analysis.' *Brit. J. Psych.: Stat. Sect.*, I, 73-81.
3. Burt, Cyril (1940). *The Factors of the Mind*. London: University of London Press.
4. Burt, Cyril (1937). 'Correlations between persons.' *Brit. J. Psych.*, XXVIII, 29-96.
5. Eysenck, H. J. (1948). *Dimensions of Personality*. London: Kegan Paul.
6. Freud, S. (1924). *Collected Papers*. London: Hogarth Press.
7. Guilford, J. P., and Guilford, R. B. (1934). 'An analysis of the factors in a typical test of introversion-extraversion.' *J. Abn. Soc. Psych.*, XXVIII, 377-399.
8. Holzinger, K. J. *Factor Analysis*. Chicago: University Press.
9. Jones, E. (1949). *Papers on Psychoanalysis*. London: Hogarth Press.
10. Kretschmer, E. (1925). *Physique and Character*. London: Kegan Paul.
11. Reyburn, H. A., and Taylor, J. G. (1943). 'On the interpretation of common factors.' *Psychometrika*, VIII, 53-64.
12. Sadger, I. (1910). 'Analerotik und analcharakter.' *Die heilkunde*, VII, 35-58.
13. Sandler, J. J. (1948). 'A factor analysis of the Rorschach test with adult mental patients.' *Proc. 12th Internat. Cong. Psych.* (in the press).
14. Stephenson, W. (1936). 'The inversed factor technique.' *Brit. J. Psych.*, XXVI, 344-61.
15. Thomson, G. H. (1946). *The Factorial Analysis of Human Ability*. London: University of London Press.
16. Thurstone, L. L. (1947). *Multiple Factor Analysis*. Chicago: University Press.
17. Weatherburn, C. E. (1947). *First Course in Mathematical Statistics*. Cambridge: University Press.

BOOK REVIEWS

Personnel Selection in the British Forces. By P. E. Vernon and J. B. Parry. London : University of London Press, Ltd. 1949. Pp. 324. 20s.

Dr. Vernon occupies a unique position among Government psychologists in this country. It is one that enables him to keep his eye on developments in four important Departments—the Admiralty, the War Office, the Air Ministry, and the Civil Service Commission. He is thus familiar with a variety of problems confronting applied psychology both in war and in peace.

Dr. Parry's work has been mainly with the Air Ministry ; and his penetrating, well-written contribution to this volume is manifestly somewhat shorter than Dr. Vernon's. Although Dr. Vernon ranges widely, it may be taken that he concentrates on work with which he himself has been fairly directly concerned. Those who are already familiar with the famous 'P.E.V.' reports, prepared during the war for the Senior Psychologist to the Admiralty and the Director of the Selection of Personnel at the War Office, will see in his chapters a much-needed summary of the most important of those reports.

Inevitably there are gaps in the narrative. Some are substantial enough to raise doubts about its apparently all-embracing title. The extremely important spade-work done by Dr. G. R. Hargreaves, particularly in the days before the formation of the Army Directorate of the Selection of Personnel, surely deserves more than the mere mention it receives on p. 40 ; and it is surprising to find the Adjutant-General's Advisory Committee dismissed with a passing comment on the same page. This group of eminent psychologists undoubtedly performed (and, indeed, still performs) a valuable function, not least in providing high-level external support for the judgments of the Army's own psychologists. Moreover, one or two of the Committee's members did in fact do very much more research on personnel selection, both for the Army and for the other Services, than might be suspected from the brief references here given.

Very little is said about statistical techniques, for (according to the preface) it is intended that information about these should be published elsewhere. It is clear, however, from other evidence, that in his own work for the Fighting Services Dr. Vernon has given full rein to his capacity for statistical improvisation, and to his taste for approximation where it seems apparent that approximation will serve. But his enthusiasm for simplicity in calculation is scarcely matched by a similar degree of enthusiasm for simplicity in the expression of results. Unlike Dr. Parry, he is not inclined to present data in diagrammatic form, despite the fact that, for at least some of the people for whom the book is intended, this is probably very desirable.

But the volume is packed with meat enough to feed, for years, psychologists and students of psychology who are interested in personnel selection and vocational guidance. Whether it will, as the authors hope, serve the purposes of the industrial personnel officer who wants to learn about the war experience of the Fighting Services is, in the reviewer's opinion, another matter. Probably there are not many readers without psychological training who will be able satisfactorily to extract what they want from so concentrated a presentation. They will be assisted by the admirable abstracts placed (very sensibly) at the beginnings of the chapters ; and perhaps would do best to start with the excellent ten-page chapter labelled 'Conclusions' and to work backwards.

Alec Rodger

The Trend of Scottish Intelligence. Sponsored by the Population Investigation Committee and the Scottish Council for Research in Education. London : University of London Press, Ltd. 1949. Pp. xxviii + 152. 7s. 6d. net.

This volume, as its sub-title indicates, gives a "comparison of the 1947 and 1932 surveys of the intelligence of eleven-year-old pupils" in Scottish schools. It will be remembered that the Royal Commission on Population was led, in the course of its inquiries, to give special attention to the possible effects of differences in the birthrate upon the general trend of national intelligence. Psychological surveys of London schools, starting as far back as those of Burt in 1913, had shown a negative correlation between size of family and tested intelligence. On the other hand, repetitions of such surveys demonstrated no very obvious decline in children's intelligence; and the general opinion of teachers and officials with long practical experience in the educational world was that, on the whole, the level of ability had shown little change during their lifetime. The negative correlation, however, was fully confirmed in other areas of the country; and certain writers, Dr. Raymond Cattell for example, gave figures for the calculated decline amounting to as much as 3 I.Q. points per generation.

In 1932 the Scottish Research Council under the Chairmanship of Mr. Hepburn (Director of Education, Ayrshire) organized a survey of all the children in Scotland born in 1921, numbering 87,498 in all. A group test was employed, and the mean score was found to be 34.4 out of a possible total of 76. Unfortunately no record was obtained in regard to size of family. In 1947, at the suggestion of the Population Investigation Committee, a similar survey was carried out with the same group test; and the mean score was found to have risen to 36.7. The increase of 2.3 points (or somewhat more on an I.Q. scale) is over 28 times its standard error. The preface discusses various explanations for this somewhat surprising result; and concludes that "we are still in the field of speculation." Certainly, the tentative inference drawn by earlier writers is by no means disproved; and it still remains possible that the level of intelligence is declining at a rate which might manifest serious effects, if sustained over a long period (longer, for example, than the interval between these surveys); for, as Professor Thomson points out, increasing familiarity with tests of the type employed may easily have masked a small actual decline.

A supplementary study was made by examining a small sample (about 1000) on each occasion with the Binet tests applied individually. The Stanford Revision was used in the first survey, and the Terman-Merrill in the second. The I.Q.'s obtained with the latter have been converted into "Stanford I.Q.'s" (on a basis of 89 cases only, however); and the correction here appears to eliminate the apparent rise in I.Q.

It is emphasized that the present volume, written by eight contributors, constitutes only a preliminary report. Additional light on the main issue may therefore be expected from the further analysis of the data—particularly from the 'sociological schedule.' Meanwhile, the account of the planning, the administrative work, the selection of tests, questionnaire-items, and the like will furnish much valuable assistance to those contemplating future surveys. The discussion of the methods used for coding data for the Hollerith counter-sorter, for checking such operations, and for detecting errors, forms a novel feature of special interest to statistical psychologists. One would have welcomed a little more information as to the nature and extent of the errors encountered; and, since the volume is a compilation rather than a narrative, an index, like that to the earlier volume, would have been most helpful.

Arthur Summerfield

INDEX TO VOLUMES I AND II

INDEX OF AUTHORS

<i>Authors</i>	<i>Titles</i>	<i>Vol.</i>	<i>Page</i>
Anstey, E.	The <i>D</i> Method of Item Analysis	I	167-177
Banks, Charlotte	Flying Ability and Body Build	I	107-113
Banks, Charlotte	Primary Personality Factors in Women : A Re-Analysis.	II	204-218
Banks, Charlotte	Factor Analysis of Assessments for Army Recruits . . .	II	76-89
Bartlett, M. S.	Internal and External Factor Analysis	I	73-81
Burt, Cyril	A Comparison of Factor Analysis and Analysis of Variance	I	3-26
Burt, Cyril	Factor Analysis and Canonical Correlations	I	95-106
Burt, Cyril	The Factorial Study of Temperamental Traits.	I	178-203
Burt, Cyril	Subdivided Factors	II	41-63
Burt, Cyril	Alternative Methods of Factor Analysis and their Relations to Pearson's Method of 'Principal Axes'	II	98-121
Burt, Cyril	The Two-Factor Theory	II	151-179
Cattell, Raymond B.	The Primary Personality Factors in Women compared with those in Men	I	114-130
Cattell, Raymond B.	A Note on Factor Invariance and the Identification of Factors	II	134-139
Cole, Raymonde	An Item-Analysis of the Terman-Merrill Revision of the Binet Tests	I	137-151
Emmett, W. G.	Evidence of a Space Factor at 11 Plus and Earlier . . .	II	3-16
Emmett, W. G.	Factor Analysis by Lawley's Method of Maximum Likelihood	II	90-97
Greenall, P. D.	The Concept of Equivalent Scores in Similar Tests . . .	II	30-40
Keir, Gertrude	The Progressive Matrices as Applied to School Children .	II	140-150
Peel, E. A.	Prediction of a Complex Criterion and Battery Reliability	I	84-94
Peel, E. A.	Item-Difficulty as the Measuring Device in Objective Mental Tests	II	69-75
Raath, M. J., and Reyburn, H. A.	Simple Structure : A Critical Examination	II	125-133
Rao, C. R., and Slater, P.	Multivariate Analysis applied to Differences between Neurotic Groups	II	17-29
Reyburn, H. A., and Raath, M. J.	Simple Structure : A Critical Examination	II	125-133
Roberts, J. Fraser	Birth Order, Maternal Age, and Intelligence	I	35-51
Sandler, J.	The Reciprocity Principle as an Aid to Factor Analysis .	II	177-184
Slater, P.	The Association between Age and Score in the Progressive Matrices Test	I	64-69
Slater, P., and Rao, C. R.	Multivariate Analysis applied to Differences between Neurotic Groups	II	17-29
Thomson, Godfrey	The Maximum Correlation of Two Weighted Batteries .	I	27-34
Thomson, Godfrey	Note on the Relations of Two Weighted Batteries . . .	I	82-83
Thomson, Godfrey	On Estimating Oblique Factors	II	1-2
Vernon, P. E.	The Variations of Intelligence with Occupation, Age and Locality	I	52-63

INDEX OF SUBJECTS

<i>Titles</i>	<i>Authors</i>	<i>Vol.</i>	<i>Page</i>
Alternative Methods of Factor Analysis and their Relations to Pearson's Method of 'Principal Axes'	Burt, Cyril	II	98-121
Analysis of Variance, A Comparison of Factor Analysis and	Burt, Cyril	I	3-26
Army Recruits, Factor Analysis of Assessments for .	Banks, Charlotte . . .	II	76-89
Battery Reliability, Prediction of a Complex Criterion and	Peel, E. A.	I	84-94

Index to Volumes I and II

Body Build, Flying Ability and	Banks, Charlotte . . .	I	107-113
Canonical Correlations, Factor Analysis and	Burt, Cyril	I	95-106
Equivalent Scores in Similar Tests, The Concept of.	Greenall, P. D.	II	30-40
Factor Analysis, Internal and External	Bartlett, M. S.	I	73-81
Factor Invariance and the Identification of Factors, A Note on	Cattell, Raymond B. . .	II	134-139
Factors, Subdivided	Burt, Cyril	II	41-63
Intelligence, Birth Order, Maternal Age, and	Roberts, J. Fraser . . .	I	35-51
Intelligence, The Variations of, with Occupation, Age and Locality	Vernon, P. E.	I	52-63
Item Analysis of the Terman-Merrill Revision of the Binet Tests, An	Cole, Raymonde	I	137-151
Item Analysis, The <i>D</i> Method of	Anstey, E.	I	167-177
Item-Difficulty as the Measuring Device in Objective Mental Tests	Peel, E. A.	II	69-75
Lawley's Method of Maximum Likelihood, Factor Analysis by	Emmett, W. G.	II	90-97
Multivariate Analysis applied to Differences between Neurotic Groups	Rao, C. R., and Slater, P.	II	17-29
Oblique Factors, On Estimating	Thomson, Godfrey . . .	II	1-2
Personality Factors, Primary, in Women : A Re-Analysis	Banks, Charlotte	II	204-218
Personality Factors, Primary, in Women compared with those in Men	Cattell, Raymond B. . .	I	114-130
Progressive Matrices Test, The Association between Age and Score in	Slater, P.	I	64-69
Progressive Matrices, The, as Applied to School Children	Keir, Gertrude	II	140-150
Reciprocity Principle, The, as an Aid to Factor Analysis.	Sandler, J.	II	177-184
Simple Structure : A Critical Examination	Raath, M. J., and Reyburn, H. A.	II	125-133
Space Factor at 11 Plus and Earlier, Evidence of a	Emmett, W. G.	II	3-16
Temperamental Traits, The Factorial Study of	Burt, Cyril	I	178-203
Two-Factor Theory, The	Burt, Cyril	II	140-166
Two Weighted Batteries, The Maximum Correlation of	Thomson, Godfrey . . .	I	27-34
Two Weighted Batteries, Note on the Relations of	Thomson, Godfrey . . .	I	82-83

BOOKS REVIEWED

<i>Author</i>	<i>Title</i>	<i>Vol.</i>	<i>Page</i>
Cattell, R. B.	Description and Measurement of Personality	I	134-136
Cattell, R. B.	A Guide to Mental Testing	II	68
Eysenck, H. J.	Dimensions of Personality	I	131-132
Finney, D. J.	Probit Analysis	I	71-72
Kelley, T. L.	Fundamentals of Statistics	I	133
Kendall, M. G.	Rank Correlation Methods	II	67-68
Scottish Council for Research in Education	The Trend of Scottish Intelligence	II	189
Thurstone, L. L.	Multiple Factor Analysis	I	70-71
Vernon, P. E., and Parry, J. B.	Personnel Selection in the British Forces	II	188
Wiener, N.	Cybernetics	II	122-124

15 Nov
T+2 in
Vol 2. Pt 2

Vol. I Part I

October 1947

THE BRITISH JOURNAL OF PSYCHOLOGY

Statistical Section

EDITED BY
CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF

A. C. AITKEN	L. S. PENROSE
M. S. BARTLETT	J. FRASER ROBERTS
W. G. EMMETT	A. RODGER
M. G. KENDALL	B. BABINGTON SMITH
D. N. LAWLEY	W. STEPHENSON
P. E. VERNON	

Managing Editor : D. W. HARDING

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4

Price Fifteen Shillings

CONTENTS

	<i>Page</i>
EDITORIAL	1
BURT, CYRIL A Comparison of Factor Analysis and Analysis of Variance	3
THOMSON, GODFREY The Maximum Correlation of Two Weighted Batteries	27
ROBERTS, J. FRASER Birth Order, Maternal Age and Intelligence	35
VERNON, P. E. The Variations of Intelligence with Occupation, Age, and Locality	52
SLATER, P. The Association between Age and Score in the Progressive Matrices Test	64
BOOK REVIEWS	70

*The British Journal of Psychology*¹ (*Statistical Section*) is issued by the British Psychological Society. Numbers will appear at intervals as sufficient material is accumulated. An effort will be made to publish at least two numbers a year.

The subscription price will be 30s. a volume of about 160 pages. Subscriptions should be sent to any Bookseller, or to THE MANAGER, University of London Press, Ltd., Warwick Square, London, E.C.4.

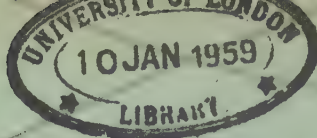
Papers for publication should be sent either to Professor Sir CYRIL BURT, Psychological Department, University College, London, W.C.1, or to Professor GODFREY THOMSON, Moray House, Edinburgh, 8. Contributors will receive 20 copies of their papers free : extra copies may be had at cost price, if the order is given when the proofs are returned.

Authors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the manuscript (including a brief list of contents at the head and a summary of results and conclusions at the end) should follow the general arrangement observed in the publications of the British Psychological Society.

Books for review should be sent to one of the Editors ; advertisements for insertion to the Manager, University of London Press, Ltd.

The British Psychological Society also issues quarterly *The British Journal of Psychology, General Section*, and *The British Journal of Medical Psychology*. Subscriptions, 30s. per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society also publishes, jointly with the Training College Association, *The British Journal of Educational Psychology*. Subscriptions for this Journal should be sent to the Managing Editor of *The British Journal of Educational Psychology*, The University, Edmund Street, Birmingham.

Papers for publication in the *General Section* should be sent to Professor F. C. BARTLETT, the Psychological Laboratory, University of Cambridge ; those for publication in *The British Journal of Medical Psychology* should be sent to Dr. JOHN RICKMAN, 3 Berkeley Court, Baker Street, London, N.W.1 ; those for publication in *The British Journal of Educational Psychology* to Professor C. W. VALENTINE, The University, Edmund Street, Birmingham.



THE YOUNG DELINQUENT

SIR CYRIL BURT, M.A., D.Sc., LL.D.

Professor of Psychology, University of London

This book analyses the chief causes of delinquency, whether arising from home circumstances, hereditary taint, physical defect and ill-health, or defects and disturbances of intelligence, temperament and character.

"A model of scientific planning . . . Nothing is irrelevant, nothing is left to chance"—*Schoolmaster*

New Impression in preparation, probably 21/- net.

THE MEASUREMENT OF ABILITIES

PHILIP E. VERNON, M.A., Ph.D.

Lecturer in Psychology, University of Glasgow

Primarily written for teachers, clinic workers, school medical officers and psychology students, it provides an elementary introduction to the statistical techniques which are essential to a proper understanding of mental measurement, and shows the application of these techniques to problems of testing and examining in English and Scottish Schools and Universities.

New Edition 15/- net.

Some Recent Work in FACTORIAL ANALYSIS and a Retrospect

GODFREY H. THOMSON, D.C.L., D.Sc., Ph.D.

Professor of Education, University of Edinburgh

This pamphlet can be recommended to those who, although having no intention of undertaking research, are anxious to know what the procedure implies. Here an intricate statistical technique is made intelligible to the average reader.

1/- net.

UNIVERSITY OF LONDON PRESS LTD

Warwick Square

London, E.C.4.

